VOL. 44 1953

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

478Z59

BOARD OF EDITORS*

STEPHEN M. COREY, *Chairman*
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

WM. CLARK TROW
University of Michigan

*Appointed by the American Psychological Association

EDITORIAL CONSULTANTS

JACK W. DUNLAP
KARL J. HOLTZINGER
HAROLD E. JONES

H. H. REMMERS
PERCIVAL M. SYMONDS

MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY

Baltimore, Md.
WARWICK and YORK, Inc.
1953

TABLE OF CONTENTS

Blood Picture in Reading, The. THOMAS H. EAMES.... 372–375

Book Reviews. 57–64, 126–128, 185–192, 249–256, 312–320, 376–382, 442–444

Comments on the Correlational Analysis Reported in "Intelligence and Cultural Differences". FRED TYLER.... 288–295

Comparison between Children Who Have Moved from School to School and Those Who Have Been in Continuous Residence on Various Factors of Adjustment, A. N. M. DOWNIE.... 50–53

Comparison of Psychology Instructors and National Norms on the Purdue Rating Scale. A. W. BENDIG 436–441

Communication Theory as the Unifying Theme in Teaching Educational Psychology. SAMUEL Z. KLAUSNER 483–489

Concept Formation of Work-study Skills by Use of Autobiographies in Grade Four. WALLACE J. HOWELL.. 257–265

Concepts of Organismic Growth: A Critique. FRED T. TYLER.... 321–342

Contributions of Film Instructions and Film Summaries to Learning from Instructional Films, The. C. W. LATHROP, JR., C. A. NORFORD AND L. P. GREENHILL 343–352

Development of a Test to Measure the Intensity of Values, The. JOSEPH E. SHORE.... 266–274

Developmental Changes in the Meaning of Minority Group Membership. MARIAN RADKE-YARROW.... 82–101

Effects of Directions Regarding Guessing on Item Statistics of a Multiple-choice Vocabulary Test. FRANCES SWINEFORD AND PETER M. MILLER.... 129–139

Effects of Practicing a Complex Arithmetical Skill upon Proficiency in Its Constituent Skills, The. WILLIAM A. BROWNELL.... 65–81

Eleven-year-old Boys in Trouble. WILLIAM W. WATTENBERG.... 410–418

Evaluation of Small Group Work in a Large Class, An. GOODWIN WATSON.... 385–409

Experiment in Evaluation in Biological Science, An. JOHN M. MASON AND GEORGE W. ANGELL.... 296–304

Experimental Comparison of the Stability of Social Acceptability Scores Obtained with the Partial-rank-

order and the Paired-comparison Scales, An. SAM L. WITHYOL AND GEORGE C. THOMPSON........... 20–30
Group Integration in a Case Discussion Course. LEONARD A. OSTLUND........................... 463–474
Group Therapy and Retarded Readers. BERNARD FISHER................................. 354–360
Improvements in Reading Rate and Comprehension of Subjects Training with the Tachistoscope. HENRY P. SMITH AND THEODORE R. TATE................ 176–184
Inter-grade Comparisons of Word Frequencies in Children's Writing. GERTRUDE HILDRETH........... 429–435
Interrelationships among Materials Read, Written and Spoken by Pupils of the Fifth and Sixth Grades. HENRY R. FEA......................... 159–175
Learning—What Kind of Animal. H. H. REMMERS.... 41–49
Mental Health Analysis of Socially Over-accepted, Socially Under-accepted, Overage and Underage Pupils in the Sixth Grade. VEGHARSH HAGOP BEDOIAN... 366–371
On the Design of Grouping Problems and Related Intelligence Tests. KARL MENGER..................... 275–287
Perceptual Retardation in Reading Disability Cases. JAMES C. COLEMAN......................... 497–503
Preferences in Colors and Illustrations of Elementary-school Children of Puerto Rico. ISMAEL RODRIQUEZ BOU AND DAVID CRUZ LOPEZ.................. 490–496
Publications Received...................... 383–384, 445–448
Qualification Responses Used with Paired Statements to Measure Attitudes toward Education. CHARLES O. NEIDT AND LYLE D. EDMISON................ 305–311
Qualitative Differences in the Vocabulary Choice of Children as Revealed in a Multiple-choice Test. LORRAINE P. KRUGLOV.......................... 229–243
Reading for Problem-solving in Science. J. HARLAN SHORES AND J. L. SAUPE..................... 149–158
Relation among Vocabulary, Structural Analysis, and Reading. JACOB TATE HUNT.................. 193–202
Relationship between Rorschach Performance and Student-teaching. MARTHA HESSEL PAGE AND ROBERT M. W. TRAVERS.............................. 31–40
Relative Contributions of Aptitude and Work Habits to

Achievement in College Mathematics. WILLIAM C. KRATHWOHL 140–148
Relative Intelligibility of Male and Female Talkers, The. B. SILVERSTEIN, R. C. BILGER, T. D. HANLEY AND M. D. STEER 419–428
Simplified Item Analysis Card for High-school and College Instructors, A. WILLIAM C. BUDD 54–56
Social Attraction Patterns between Elementary-school Children and Student-teachers: Sociometric Analysis. WALTER D. SMITH 113–125
Some Effects of Bilingualism upon the Intelligence Test Performance of Puerto Rican Children in New York City. ANNE ANASTASI AND FERNANDO A. CORDOVA . . . . 1 –19
Some Relationships between Non-intellectual Characteristics and Academic Achievement. JOHN P. MCQUARY 215–228
Some Relationships between Students' Perceptions of School and Their Achievement. LESLIE F. MALPASS 475–482
Study of Social Sensitivity (Sympathy) among Adolescents, A. WALTER LOBAN 102–112
Teacher Growth in Attitudes toward Behavior Problems of Children. MANFRED H. SCHRUPP AND CLAYTON M. GJERDE 203–214
"The Essential High School Content Battery" as a Predictor of College Success. MARIE P. DOLANSKY . . . 361–365
Validity of the Goodenough Draw-a-Man Test in Greece, The. I. TH. PAPAVASSILIOU 244–248
Value Concept in Educational Psychology, The. WM. CLARK TROW 449–462

BOOK REVIEWS

Dorothy C. Adkins and Samuel B. Lyerly. *Factor Analysis of Reading Tests.* (EDWARD E. CURETON) 189
Samuel J. Beck. *Rorschach's Test II—Advances in Interpretation.* (DAVID WECHSLER) 378

Henry Bowers. *Research in the Training of Teachers.* (H. H. Remmers) 312

Daniel Brower and Lawrence E. Abt, Editors. *Progress in Clinical Psychology.* (A. S. Edwards) 376

Raymond B. Cattell. *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist.* (W. J. E. Crissy) 254

James Deese. *The Psychology of Learning.* (Miles A. Tinker) 382

Horace B. English. *Child Psychology.* (Melvin G. Rigg) 128

Miriam Forster Fiedler. *Deaf Children in a Hearing World: Their Education and Adjustment.* (Eloise Kennedy) 441

John W. French. *The Description of Aptitude and Achievement Tests in Terms of Rotated Factors.* (Lee J. Cronbach) 249

Milton Gurvitz. *The Dynamics of Psychological Testing.* (Lee J. Cronbach) 185

Ruth E. Hartley, Lawrence K. Frank, and Robert M. Goldenson. *Understanding Children's Play.* (R. Elizabeth Brown) 191

Higher Education in the Forty-eight States: A Report to the Governors' Conference. (C. M. Louttit) 440

Lloyd J. Jeffress. *Cerebral Mechanisms in Behavior.* (H. Meltzer) 442

Louis Kaplan and Denis Baron. *Mental Hygiene and Life.* (A. S. Edwards) 61

G. Marian Kinget. *The Drawing-completion Test.* (Goldine C. Gleser) 251

Margaret Mead and Frances Cooke MacGregor. *Growth and Culture: A Photographic Study of Balinese Childhood.* (Frank Costin) 316

William C. Morse, Francis A. Ballantine, and W. Robert Dixon. *Studies in the Psychology of Reading.* (Miles A. Tinker) 188

James L. Mursell. *Using Your Mind Effectively.* (Philip M. Kitay) 318

Margaret Naumburg. *Schizophrenic Art: Its Meaning in Psychotherapy.* (H. Meltzer) 63

C. W. Odell. *How to Improve Classroom Testing.* (Chester W. Harris) 320

Otto Pollack et al. *Social Science and Psychotherapy for Children.* (A. S. Edwards) 314
Harold G. Shane and E. T. McSwain. *Evaluation and the Elementary Curriculum.* (Arthur W. Foshay) 57
G. Milton Smith. *More Power to Your Mind.* (Miles A. Tinker) 315
Calvin P. Stone, Editor. *Annual Review of Psychology.* Volume 3. (Miles A. Tinker) 249
Edward S. Strecker, Franklin C. Ebaugh, Jack R. Ewalt, and Leo Kanner. *Practical Clinical Psychiatry.* (H. Meltzer) 126
George G. Thompson. *Child Psychology: Growth Trends in Psychological Adjustment.* (C. M. Louttit) 187
Arthur E. Traxler, Robert Jacobs, Margaret Selover, and Agatha Townsend. *Introduction to Testing and the Use of Test Results in Public Schools.* (Dora A. Damrin) 381
Dael Wolfle, Claude E. Buxton, Charles N. Cofer, John W. Gustad, Robert B. MacLeod, and Wilbert J. McKeachie. *Improving Undergraduate Instruction in Psychology.* (Miles A. Tinker) 380
Dorothy Yost Deegan. *The Sterotype of the Single Woman in American Novels.* (Miles A. Tinker) 59

INDEX TO AUTHORS

Anastasi, Anne	1
Angell, George W.	296
Bedoian, Vagharsh Hagop	366
Bendig, A. W.	436
Bilger, R. C.	419
Bou, Ismael Rodriquez	490
Brownell, William A.	65
Budd, William C.	54
Coleman, James C.	497
Cordova, Fernando A.	1
Dolansky, Marie P.	361
Downie, N. M.	50
Eames, Thomas H.	372
Edmison, Lyle D.	305
Fea, Henry R.	159
Fisher, Bernard	354
Gjerde, Clayton M.	203
Greenhill, L. P.	343
Hanley, T. D.	419
Hildreth, Gertrude	429
Howell, Wallace J.	257
Hunt, Jacob Tate	193
Klausner, Samuel Z.	483
Krathwohl, William C.	140
Kruglov, Lorraine P.	229
Lathrop, C. W., Jr.	343
Loban, Walter	102
Lopez, David Cruz	490
Malpass, Leslie F.	475
Menger, Karl	275

Mason, John M. 296
McQuary, John P. 215
Miller, Peter M. 129
Neidt, Charles O. 305
Norford, C. A. 343
Ostlund, Leonard A. 463
Page, Martha Hessel 31
Papavassiliou, I. Th. 244
Radke-Yarrow, Marian 82
Remmers, H. H. 41
Saupe, J. L. 149
Schrupp, Manfred H. 203
Shore, Joseph E. 266
Shores, J. Harlan 149
Silverstein, B. 419
Smith, Henry P. 176
Smith, Walter D 113
Steer, M. D. 419
Swineford, Frances 129
Tate, Theodore R. 176
Thompson, George C. 20
Travers, Robert M. W. 31
Trow, Wm. Clark 449
Tyler, Fred T. 288, 321
Watson, Goodwin 385
Wattenberg, William W. 410
Withyol, Sam L. 20

No. 1 Missing
PNB

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 44 February, 1953 Number 2

478 X 5

THE EFFECTS OF PRACTICING A COMPLEX ARITHMETICAL SKILL UPON PROFICIENCY IN ITS CONSTITUENT SKILLS[1]

WILLIAM A. BROWNELL
University of California

```
      79, R22
34/2708
   238
   ---
    328
    306
    ---
     22
```

Success in completing division in examples like that at the right depends upon knowledge of the basic combinations in the four fundamental operations and also upon

1) proficiency in simple multiplication (7 × 34; 9 × 34), including carrying;

2) proficiency in subtraction, without and with 'borrowing' (328 − 306; 270 − 238); and

3) proficiency in simple division (as in $3/\overline{270}$), to the extent of knowing the process and the algorism (method of setting down numbers for computation).

Division by two-place divisors, as in the type example, is taught after skills (1), (2), and (3), the so-called long form now being generally used from the outset.

It can be argued that, before undertaking the more complex skill of dividing by two-place numbers, children should be letter-perfect in its subordinate skills as listed above, or in its sub-skills as they will be designated hereafter. Actually under the conditions of ordinary classroom teaching such mastery is rare. Instead, children start to learn the complex skill with varying degrees of competence and incompetence in simple multiplication, subtraction, and division.

CONFLICTING HYPOTHESES

On this account some would advocate deferring division by two-place divisors, until after proficiency in the sub-skills has been developed. Their argument is that children deficient in the sub-skills must find the more intricate skill too much for them; they must be confused by its demands and may lose control, at least temporarily, over the sub-skills themselves. In their view it is false economy to move ahead too fast.

---

[1] It is a pleasure to acknowledge the financial assistance of Northwestern University in making this study possible.

Others take a contrary position. They are ready to start teaching the complex skill without a high degree of competence in the sub-skills. Their case is somewhat as follows: Practice on a total skill brings improvement in its constituent skills, since the latter are then learned in their functional relations.

DISCUSSION

There is, of course, evidence of a kind for both opposing positions. Children as they move upward in the grades do gain greater proficiency in skills taught them earlier. For example, when they are in Grade 4, children score higher on Grade 3 arithmetic tests than they did the year before. Just how the improvement in the test scores is to be explained is another matter. In the first place, as claimed, it may occur more or less incidentally as the previously learned skills are put to work in more complex relationships. Or, in the second place, it may result from a quite different cause: namely, remedial instruction directed to the removal of specific learning shortages. Or, in the third place, it may follow from the operation of both sets of conditions in varying combinations.

At least one research report seems to support the practice of moving ahead even when prior learning is incomplete. Breed and Ralston[2] taught the simple addition combinations to two groups of second-grade children. All subjects studied the combinations as such for a time. One group continued under the usual kinds of drill. The experimental group, however, turned to column addition in which they used the simple combinations. On a test on the combinations given at the close of the investigation, the latter children were clearly superior. Just why, however, is not so clear. During the time they were at work on column addition, the combinations with their answers were on

---

[2] Frederick S. Breed and Alice L. Ralston, "Direct and indirect methods of teaching the addition combinations." *Elementary School Journal*, 37: 283–94, December, 1936.

478 X 57

display for ready reference. Was it this fact, or was it the practice on the more complex skill, that produced their greater gains? Or, was it both?

To turn to the other side of the controversy: there is some research support for the position that 'haste makes waste.' Swenson[3] taught the addition combinations in a series of groups or sets. She reports that facts learned later sometimes had a harmful backward effect upon facts learned earlier. The amount of this effect, known as retroactive inhibition, varied with a number of factors. For example, interference was less, the more meaningful the learning. The Swenson study of course dealt with a problem different from that in the present investigation and employed number facts rather than a hierarchy of skills; but it is relevant here since it shows unmistakably that what is learned at any one time in arithmetic may impair what was learned before then.

Another study,[4] dealing with skills, demonstrates the reality of retroactive inhibition in the case of subtraction. Before starting to learn how to subtract in examples involving 'borrowing' (e.g., 71 − 43), the third-grade children who served as subjects were tested for proficiency in 'non-borrowing' examples (e.g., 87 − 53). A test, including examples of the 'non-borrowing' type and given after two weeks of study devoted to 'borrowing' subtraction, disclosed that many of the children had lost competence in the more familiar kind of subtraction—had even lost command of many simple number combinations.

THE PRESENT INVESTIGATION: THE TESTING

The research findings summarized in this brief review are certainly inconclusive in their implications. They warrant no confident prediction as to the consequences, helpful or harmful, of teaching division by two-place divisors to children with imperfect mastery of the prerequisite subskills. The present study was designed to throw light on the matter.

---

[3] Esther J. Swenson, "Organization and generalization as factors in learning, transfer, and retroactive inhibition," in *Learning Theory in School Situations.* University of Minnesota Studies in Education, College of Education, No. 2. Minneapolis: The University of Minnesota Press, 1949, pp. 40–74.

[4] William A. Brownell, et al. *Learning as Reorganization.* Duke University Research Studies in Education, No. 3. Durham, N. C.: Duke University Press, 1939.

A Readiness Test for division by two-place numbers was administered to the children in seventeen fifth-grade classes in the schools of Cicero, Evanston, and Waukegan, Illinois. Instruction in the new, more complex skill was to be begun at once. The test, a copy of which is presented, consists of twenty-eight examples, seven each in subtraction (with and without 'borrowing') and in multiplication (with and without 'carrying') and fourteen in division by digits. The children worked the examples across the page in cycles—one in subtraction, then one in multiplication, next two in division, and then back to subtraction again. Since the purpose was to have scores on the test as a whole, no time limit was set. Only accuracy scores, therefore, appear in the results here reported.

After three weeks of study on division with two-place divisors, the Readiness Test was again given the children under conditions like those in the first administration. Two test papers were thus made available for a population of 367 children. The findings below are based upon comparisons of the gross scores and of partial scores on the two tests. It should be added that the reliability of the test was found to be $r_{11} = .91$ (split-halves method corrected by the Spearman-Brown formula).

GROSS FINDINGS

Table 1 is the master exhibit. It is read as follows: On the first test (vertical scale) twenty-two children made perfect scores. Of this number, one made a score of 24 on the second test (horizontal scale), four a score of 25, three a score of 26, and four a score of 27, while ten repeated with perfect scores. The heavily marked series of cells rising diagonally from the lower left corner to the top right corner contain the numbers of children who made identical scores on the two tests. Numbers in cells to the right of this axis refer to children who improved their scores, and those to the left of the axis to those who made poorer scores the second time. Yet, these statements are obvious over-simplifications, for the scores cannot be accepted at face value. Both variable and systematic factors must be taken into account. The $PE_{meas.}$ on the pre-test was 1.52. Changes of one point in scores are therefore best regarded as unreliable. Gains or losses of 2 points,

TABLE 1.—MASTER EXHIBIT OF GROSS FINDINGS

SECOND TEST

FIRST TEST	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	N
28																									1	4	3	4	10	22
27															1								1	1	2	8	4	9	4	30
26																				1	1			6	5	7	7	10	2	39
25																			1	1		5		5	2	8	8	8	1	39
24																					1	2	1	6	11	7	8	3	3	42
23														2					1		1	1	3	6	4	9	4	4	1	36
22																			1		2	2	1	2	10	4	3	1	1	27
21															1					1	2	3	1	6	1	6	2	1		24
20															1		2	2	2		2	1	4		3	1		1		19
19																		1		2		1		2	2	1				9
18																3	1	1			2			1	1	1		2		12
17												1	1			1	1		1	3	1	1	1		1		1			13
16																1			1						1					3
15								1	1		1	1	1		2		1		1		1	2								12
14															1			1			1	1								4
13									1			1		1		1					1		2							7
12								1		1				.	1															3
11								1			3	2			1															7
10										1					1	1														3
9										1																				1
8								1																						1
7												2				1														3
6				1							1																			2
5								2																						2
4									1																					1
3												1																		1
2						1	1																							2
1					1																									1
0	2																													2
N				1	1	1	1	6	3	3	5	8	2	3	9	8	5	5	8	8	15	19	14	35	44	56	40	43	22	N = 367

as from 22 to 24 or from 24 to 22, are more reliable, representing eighty chances in a hundred of being real gains rather than losses, and vice versa.

READINESS TEST

1. $\begin{array}{r}219\\ \times 4\\ \hline\end{array}$	2. $\begin{array}{r}780\\ -69\\ \hline\end{array}$	3. $3\overline{)963}$	4. $4\overline{)849}$
5. $\begin{array}{r}129\\ \times 6\\ \hline\end{array}$	6. $\begin{array}{r}910\\ -78\\ \hline\end{array}$	7. $6\overline{)1380}$	8. $8\overline{)3449}$
9. $\begin{array}{r}897\\ \times 9\\ \hline\end{array}$	10. $\begin{array}{r}836\\ -279\\ \hline\end{array}$	11. $7\overline{)4553}$	12. $4\overline{)2930}$
13. $\begin{array}{r}470\\ \times 8\\ \hline\end{array}$	14. $\begin{array}{r}942\\ -368\\ \hline\end{array}$	15. $6\overline{)4220}$	16. $9\overline{)3829}$
17. $\begin{array}{r}654\\ \times 9\\ \hline\end{array}$	18. $\begin{array}{r}504\\ -476\\ \hline\end{array}$	19. $7\overline{)5667}$	20. $8\overline{)1300}$
21. $\begin{array}{r}968\\ \times 8\\ \hline\end{array}$	22. $\begin{array}{r}781\\ -596\\ \hline\end{array}$	23. $8\overline{)7028}$	24. $8\overline{)5268}$
25. $\begin{array}{r}689\\ \times 7\\ \hline\end{array}$	26. $\begin{array}{r}503\\ -258\\ \hline\end{array}$	27. $7\overline{)6006}$	28. $9\overline{)7000}$

There are more entries in the cells to the right of the axis than in those to its left, with a corresponding increase of two points in the second test median (from 22 to 24). Some part of this gain (its amount being unknown) is probably attributable to practice effect in taking the same test twice.

If learning a new complex skill greatly increases proficiency in its sub-skills, as is claimed, there is no substantial evidence thereof in Table 1. Of course pupils making scores of 26 or higher had little chance to improve; but those making scores of 19 or less had ample chance. There were eighty-nine such children. Fifty-one of them did improve if (as we should not) we take scores as they stand, without correction of any kind. On the other hand, twenty-nine children, or about one third of the group in question, made lower scores the second time. It is especially noteworthy that those with the greatest opportunity to move forward did not do so to any great extent. Sixteen children made scores of 9 or less on the first test. Two of these subjects gained 8 points on the second test; the other six gained 4 points or less. True, in every instance their absolute gains equalled or exceeded the average gain for the 367 children as a whole; but the educational significance of the gains is negligible. Only one subject attained an accuracy score as high as fifty-four per cent, and only two others, as high as forty per cent.[5] Clearly these pupils who were most seriously weak in the sub-skills got little help on them by learning the complex skill. There must be better ways than this to remove deficiencies in such cases.

Just as one looks to the bottom of Table 1 for evidence of positive effects from learning the complex skill, so one looks to the top of the table for evidence of negative effects. Obviously children with high initial scores had the most to be interfered with. There are numerous cases in the table where losses of a point occurred; but these had best be discarded for reasons given. On the other hand, there are instances in which something untoward must almost certainly have occurred. One child dropped in his scores from 26 to 19, probably not a matter of chance. Two who first made scores of 25 later made scores of 18 and 19. Two scored 13 after first scoring 23, and two scored 7 and 8 after scoring 15. In these cases, at least, retroactive inhibition seems to have been at work; but its effect was not uniform and general, for at each score level where there were losses on the second test, there were as many or more gains, and of comparable size.

The data in Table 1 speak unequivocally neither for negative nor for positive effects: they speak for both. Learning the complex skill of dividing by two-place numbers led some children, directly or indirectly, to improve their mastery of its constituent sub-skills; but just as truly, it led other children, for the time at least, to be less proficient therein. We have here no all-or-none proposition, as seems to be assumed. Retroactive inhibition from the complex skill may occur, or it may not. Similarly, greater mastery of sub-skills may ensue, or it may not.

---

[5] These gains can be interpreted in still another way. For the individual children concerned, they may represent truly extraordinary improvement, despite their relative insignificance in terms of acceptable educational standards of competence.

MORE DETAILED FINDINGS

1) Gains and Losses by Process

A table like Table 1 was made for each of the three fundamental operations included in the Readiness Test. Data for these tables, omitted here, were obtained from one hundred twenty-three paired test papers. Approximately each third child in the alphabetical lists for the seventeen classes was taken to constitute the sample. Interpretation of these data is admittedly a bit hazardous. Scores could range only between 0 and 7 in subtraction and in multiplication, and between 0 and 14 in division, and the significance of small differences in scores is correspondingly open to question.[6]

Subtraction.—Of the one hundred twenty-three subjects, twenty-eight (twenty-three per cent) had lower scores on the second test as compared with forty-seven (thirty-seven per cent) who had higher scores. Such losses as occurred were slight, amounting to as much as 2 points in only eight instances. Improvement was of course impossible for the forty-eight children who made perfect scores of 7 on Test 1. Of the forty-three who could have gained 2 points or more, twenty-six did so. On the other hand, where the chance for learning was maximal, there was not a great deal. Of the eleven children with initial scores of 3 (forty-two per cent accuracy) or less, only six raised their standing, and these by insubstantial margins. As a whole, the distributions revealed no general tendency for scores to change in one direction or the other. A very few cases seem to show the harmful effects of retroactive inhibition, but relatively more cases appear to show improvement in subtraction through experience with the new type of division.

Multiplication.—In multiplication thirty-nine children had higher scores, and thirty-eight lower scores on the second test—almost an exact balance. Fifteen scored 3 points or less the first time, fourteen of them maintaining or raising their scores the second time. The gains in three instances were of but a single point. However, three children who made scores of 2 originally moved to scores of 5 or 7; and one who scored a single point at first moved to 6. Learning the complex skill had positive effects then in the case of some of the children at the lower end of the scale.

---

[6] The issue here does not seem to be resolved by statistics; otherwise, $PE_{meas.}$'s would have been calculated. On a scale of scores from 0 to 7, a change of one point takes on exaggerated importance, especially when so small a change could often result from chance.

But learning to divide by two-place numbers had negative effects, too, discernible particularly among the children at the upper end of the distribution. Of forty children who made perfect scores of 7 on Test 1, eight suffered losses of 2 points on Test 2. Of thirty-three who made initial scores of 6, one pupil dropped to a score of 2, and two to scores of 3. Of twenty-one who made initial scores of 5, one dropped to 1, a loss of 4 points.

All in all, then, competence in multiplication was sometimes increased and sometimes lessened following instruction on the new division skill. The results for multiplication are, therefore, similar to those for subtraction in that both types of change occurred; but they are different, too, for in multiplication there were more instances of extreme change, particularly in the direction of lower proficiency.

Division.—On the first test, one hundred ten subjects scored 8 (fifty-seven per cent accuracy) or higher. Thirty-two of these children made lower scores on the re-test, one falling from a perfect score of 14 to 0, while thirty-eight raised their scores. Losses amounting to 2 points or more occurred in fifteen instances, and gains of 2 points or more, in twenty-two instances. For children, then, who had accuracy scores of about sixty per cent or better in simple division, the introduction of the new complex division skill seems to have had variable effects, both helpful and harmful.

For those with scores of fifty per cent accuracy or less on Test 1, the effects were more generally harmful. Three of the thirteen children in this category raised their second scores by 2 points or more, but six suffered losses of comparable size. Yet, in view of the fact that one of the thirteen subjects gained 6 points, it is impossible to say that retroactive inhibition was the necessary fate of children low in the scale of achievement in simple division.

INTERPRETATION

The foregoing analyses furnish bases to test two hypotheses, both relating to retroactive inhibition. One hypothesis is that the new complex skill should have disturbed least the sub-skills that had been mastered most thoroughly. On the first test the mean scores represented accuracies of eighty-seven per cent, seventy-eight per cent, and seventy-six per cent, respectively, in subtraction, multiplication, and division. Proficiency in subtraction less commonly suffered from interference than did proficiency in the other two operations. In this respect, then, findings in the present study are consistent with those in investigations of retroactive inhibition in general.[7]

The research on retroactive inhibition just alluded to has found that, up to a point, the closer the similarity between a new and an old learning task, the greater the amount of interference. Now, division by two-figure divisors resembles division by digits more closely than it does subtraction or multiplication taken separately. Accordingly, proficiency in simple division should have been affected adversely more than proficiency in the other two sub-skills; and this is the second hypothesis on which there are data in the present study. The hypothesis is borne out in part, better (but by no means perfectly) in the case of children low in achievement than in the case of those relatively higher in achievement. For the latter subjects evidence of improvement in simple division was fully as common as was evidence of retroactive inhibition.

2) Types of Error

Examination of papers secured in the second testing disclosed no frequently occurring types of error in subtraction or in multiplication that could be attributed to retroactive inhibition. In simple division, however, there were five common or fairly common types of error, some of which seem to be related to interference from the more complex kind of division. These five types are listed in Table 2 with their frequency, and are illustrated in the text.

The commonest error was Type 1, and like Types 2 and 3 it appeared in all three experimental centers on the second test papers but not the first. The examples with two-place divisors which the children had been solving all had had quotients of one- or of two-place numbers. One may conjecture that they had developed an expectancy of this kind of quotient and that it operated when they solved the examples in simple division in the second test.

---

[7] For an excellent and helpful analytic summary of theory and research, see: Esther J. Swenson, *Retroactive Inhibition, A Review of the Literature*. University of Minnesota Studies in Education, College of Education, No. 1. Minneapolis: The University of Minnesota Press, 1941.

TABLE 2.—TYPES OF DIVISION ERROR ON THE SECOND TEST

Type	Frequency
1) Too few figures in the quotient	129
2) Placing quotient figures incorrectly	109
3) Failure properly to bring down figures from the dividend	66
4) Obtaining complete quotient before getting partial dividend	25
5) Interchange of figures in the quotient	13

Type 2 errors consisted in misplacing quotient figures, always too far to the right. Explanation here is uncertain. It is clear that the children who made this error did so mechanically; that is, without understanding the algorism. Since this error did not appear in their first test papers, these particular children were apparently exhibiting in the second test some kind or kinds of confusion brought on by learning the new division skill.

Type 1

$$\begin{array}{r} 8\ 7 \\ 7/\overline{5667} \end{array} \qquad \begin{array}{r} 65 \\ 7/\overline{4553} \\ \underline{42} \\ 35 \\ \underline{35} \end{array}$$

Type 2

$$\begin{array}{r} 320 \\ 3/\overline{963} \end{array} \qquad \begin{array}{r} 651 \\ 4/\overline{4553} \end{array}$$

Type 3

$$\begin{array}{r} 801 \\ 7/\overline{5667} \\ \underline{56} \\ 0 \\ \underline{6} \\ 07 \\ \underline{7} \end{array} \qquad \begin{array}{r} 210 \\ 8/\overline{1380} \\ \underline{8} \\ 50 \\ \underline{48} \\ 2 \end{array}$$

Type 4

$$\begin{array}{r} 801 \\ 7/\overline{5667} \\ \underline{5607} \\ 60 \end{array} \qquad \begin{array}{r} 210 \\ 6/\overline{1380} \\ \underline{1260} \\ 20 \end{array}$$

Type 5

$$\begin{array}{r} 123 \\ 3/\overline{963} \end{array} \qquad \begin{array}{rl} 730 & \text{r} - 2 \\ 6/\overline{4220} & \\ \underline{42} & \\ 20 & \\ \underline{18} & \end{array}$$

Type 3 errors resulted from failure properly to bring down figures from the dividend. No better explanation can be offered for Type 3 errors than for Type 2 errors. All one can say is that in learning to divide by two-place divisors apparently 'something' happened that affected skill in simple division for the worse.

Type 4 and Type 5 errors occurred in the classes of only one experimental center. They may or may not have resulted from interference from the new division skill; but, if so, how the effect was produced is unclear. In the case of Type 4 errors the children seem to have started to divide by the new method and then to have changed in midstream to the old method. Observations of their work or questioning during the work, neither of which was done in this study, might have revealed reasons for the errors. Only the fact that Type 4 and Type 5 errors appeared exclusively on the second test papers warrants the belief that they were caused by retroactive inhibition.

3) Results in the Three Experimental Centers

Center 1.—As shown in Table 3 the mean scores in the two tests in Center 1 were identical, 21.8. Does this imply that no changes at all occurred in the three sub-skills, that proficiency therein was neither favorably nor unfavorably influenced by three weeks' experience in dividing by two-place numbers?

TABLE 3.—MEAN SCORES ON THE TWO TESTS IN THE THREE CENTERS

Test	Means by Centers		
	Center 1 (N = 143)	Center 2 (N = 60)	Center 3 (N = 157)
Test 1	21.8	19.0	22.3
Test 2	21.8	19.2	23.5

The tabulation in Table 4 reveals that, despite surface indications to the contrary, a great deal of change occurred in the sub-skills. The scores of ninety subjects (sixty-three per cent of the population) differed by 2 points or more on the two tests, and the differences were in both directions. For forty-four pupils the second test scores were higher, in nineteen instances by 4 points or more. (Four children gained 5 points, three

6 points, one each 7 and 8 points, and two 9 points.) On the other hand, for forty-six children the second scores were lower by 2 points or more, in eighteen instances by 5 points or more. (Two pupils lost 5 points; four, 6 points; two, 7 points; one each, 10 and 13 points.)

TABLE 4.—CENTER 1: INSTANCES OF CHANGES OF 2 OR MORE POINTS ON TEST 2

Scores, Test 1	No.	Gains, in Points			Losses, in Points		
		2	3	4, or more	2	3	4, or more
24–28	34	7	2	0	10	6	9
20–23	30	7	4	9	2	2	6
16–19	9	1	1	2	1	3	1
12–15	10	0	2	5	1	1	1
11, or less	7	1	0	3	1	1	1
Totals	90	16	9	19	15	13	18

In Center 1, then, the situation respecting the sub-skills was anything but stable, as might be inferred from a comparison of means only. The latter measures remained unchanged because improvement was off-set by retroactive inhibition, the net consequence being 0.[8] The conclusion for Center 1 is that practice on the complex skill had variable effects, making for gains in the case of some children and for losses in the case of other children, and in equal measure.

Center 2.—The mean score on Test 2 was 0.2 points higher than that on Test 1, the difference being unreliable. As in the case of

---

[8] There is an interesting parallel here with experimentation on transfer of training. When in this latter research 'zero transfer' or 'no transfer' is reported, the conclusion as stated is probably questionable, for the phrase refers properly to net consequences rather than to transfer as such. In such instances there may actually be a good deal both of positive and of negative transfer, the effects of facilitation cancelling those of interference. Rarely is the total amount of transfer reported, in which cases measures both of positive and of negative transfer would be combined, rather than subtracted, one from the other.

Center 1, the similarity of means could be interpreted as meaning that practice with two-place divisors had no effect at all upon the sub-skills, or that this practice had variable effects that balanced each other. The second interpretation is the correct one.

A table like Table 4 for Center 1 was constructed for Center 2. In the interests of brevity, this table is omitted, and the main findings are summarized as follows:

Number of children with changes of 2 points or more on Test 2	36	(60% of the population)
Number of children gaining 2 points or more...	20	
Number of children gaining 4 points or more... (Two gained 6 points; one, 7 points; one, 8 points; one, 9 points)	8	
Number of children losing 2 points or more.....	16	
Number of children losing 4 points or more..... (Two lost 5 points; two, 6 points; two, 7 points; one, 9 points)	9	

Clearly, then, in Center 2 there was improvement in the sub-skills, but there was about the same amount of deterioration. When, however, the unlike effects are brought together for the thirty-six children and are consolidated into a single figure (the average), one may draw the erroneous conclusion that practice on the complex skill had no influence upon its constituent sub-skills.

Center 3.—The results in Center 3 are unlike those in the other two centers. The one hundred fifty-seven children in this group were superior to those in the other two groups at the outset (Table 3), and they increased their advantage in Test 2. Their mean on Test 2 was 1.2 points higher than their mean on Test 1, and the difference is a reliable one (CR = 5.57). As a whole, the children in Center 3 seem to have been aided rather than harmed by experience with two-place divisors. Seventeen per cent of these children did suffer losses of 2 points or more on Test 2 (the table, like Table 4, is omitted), but this figure is to be compared with per cents of thirty-two and twenty-eight in Centers 1 and 2, respectively. Furthermore, there was only one loss as large as 5 points, and two more as large as 4 points.

There are other reasons to believe that in Center 3 retroactive inhibition did not operate as generally or as seriously as in the other centers. (a) The total number of Type 1, Type 2 and Type 3 errors attributed in Table 3 to retroactive inhibition represented an average incidence of 0.20 per pupil in Center 3 as compared with 0.92 in Center 1 and with 2.60 in Center 2. (b) An impressionistic classification revealed that twenty of the one hundred forty-three children in Center 1 appeared to be seriously confused by what they had to do in Test 2, and these twenty were distributed among all seven of the classes in that center. In Center 2, with a population of sixty, there were eleven such children, and they were enrolled in three of the five classes. In Center 3, with a population of one hundred fifty-seven, the total was only eight, seven in one class.

The unlikeness of the results in Center 3 correlates with known differences in instruction in this center. In Center 3 alone there was a systematic program of individualized teaching. Children were grouped according to their learning needs and received specialized help. As is usually the case, differentiated instruction paid off. Learning difficulties could be detected early and dealt with at once.

So far as this experiment is concerned, the effects of the remedial teaching in Center 3 are two in number. In the first place, specialized instruction greatly reduced opportunity for retroactive inhibition by removing its harmful consequences as quickly as they appeared. In the second place, it greatly reduced the credit to be accorded practice on the complex skill as the cause of improvement in the sub-skills. The gains made may have been wholly or largely produced by the remedial measures, quite apart from experience with the complex skill.

SUMMARY

In the absence of differentiated remedial instruction (as in Center 3), experience with two-place divisors had variable results (Centers 1 and 2). It was just as likely to lead to improvement in the sub-skills as to deterioration therein (and vice versa). In these circumstances gains and losses equalled each other, with the net change amounting to 0.

4) Effects upon Individual Children

Table 5 contains the records made by ten children on the two tests. Their scores are entered for the tests as wholes, as well as their sub-scores on the tests for the sub-skills. The ten subjects were selected from Centers 1 and 2 where the influence of the new division skill could be studied more directly than in Center 3, without complications from the extraneous factor of special remedial instruction.

TABLE 5.—RECORDS OF TEN SELECTED SUBJECTS

Subject	Test 1	Test 2	Subtraction		Multiplication		Division	
			1	2	1	2	1	2
A	28	28	7	7	7	7	14	14
B	13	22	1	6	2	4	10	12
C	18	27	4	6	6	7	8	14
D	25	19	5	3	7	5	13	11
E	23	13	4	3	6	2	13	8
F	15	20	0	3	4	6	11	11
G	21	26	4	7	6	6	11	13
H	13	22	3	4	3	4	7	14
I	15	21	3	4	6	4	6	13
J	20	14	6	3	4	6	11	5

About all the possible relationships between experience with the new division skill and its effects upon the sub-skills are illustrated in Table 5. Subject A, with a high degree of initial mastery of the sub-skills maintained that mastery without change, at least so far as change was here measured. Subjects B and C seem to have improved generally in the sub-skills. By contrast, Subjects D and E lost proficiency generally. Subject F gained in subtraction and multiplication without improvement in division. Subject G gained in subtraction and division, with no change in multiplication. Subject H increased greatly his score in simple division, but increased his scores in the other two sub-skills little if any. Subject I also went 'up' considerably in division, but did less well in multiplication. Subject J was equally inconsistent, improving in multiplication but losing materially in the other sub-skills.

Other cases could have been added to the table, but they would have served only to confirm the facts to be noted in those cited: learning to divide by two-place numbers had no single uniform result, good or bad. Instead, its effects varied from child to child, and in many instances from sub-skill to sub-skill for the same child.

CONCLUSIONS AND INTERPRETATION

1) Practice in dividing by two-place numbers (the complex skill) had no single, uniform and predictable result so far as proficiency in the sub-skills is concerned. In a given class the effects were both helpful and harmful, and sometimes in the same child helpful in some sub-skill and harmful in others.

2) In general, the oldest and best established sub-skill (subtraction) seemed to be less subject to change (improvement or deterioration) than sub-skills more recently taught, while the sub-skill (simple division) most like the complex skill seemed to be least stable.

3) It is safer to attribute loss in proficiency in sub-skills to retroactive inhibition from the new complex skill than it is to attribute improvement to practice on that skill. Gains may be produced by factors (such as individualized remedial instruction) which are not necessarily involved in practice on the complex skill.

4) Children with the lowest degree of proficiency in the sub-skills made relatively little improvement therein while working on the new complex skill. For such children carefully planned remedial teaching is to be recommended.

DEVELOPMENTAL CHANGES IN THE MEANING OF MINORITY GROUP MEMBERSHIP[1]

MARIAN RADKE-YARROW

Human Resources Research Office
George Washington University

In many situations and in various ways, the individual's personal and social behavior is influenced by the groups to which he belongs. He is influenced also by the standards of groups to which he does not belong, as the standards of these groups exert pressures upon him and interfere with or facilitate the attainment of his goals. The question for research is no longer, do group memberships influence the individual, but rather, how do specific group memberships affect the individual and how are these effects influenced by specific environmental and personality factors.

Membership in a cultural minority group often creates special problems for the individual. The group may provide him with various rewards and securities, but it is almost certain, also, to generate problems of frustration and uncertainty, of conflicting values and loyalties. Perceptions of the social world may be influenced greatly by needs and values stemming from the individual's minority belonging.

Certainly the child's conceptions of himself and society do not escape the impact of minority status. How early are these social tensions experienced by the child? What kinds of barriers are felt most keenly by him? What is his understanding of the social distinctions which he has 'inherited' and of the social punishments imposed upon him? How does he achieve integration into the general culture of his school and community?

Answers to these questions would be of great help to educators who are attempting to meet the special social needs of minority children. Efforts at improving minority-majority group relations and minority group morale might be directed more effectively by an understanding of the development of psychological minority status within the individual. Those who are concerned with the social order, who are striving to perpetuate and to realize more fully democratic ideology and democratic living, can not fail to be concerned about the effects of social inequalities upon the developing individual.

---

[1] This study was done for the Commission on Community Interrelations of the American Jewish Congress, 1834 Broadway, New York City.

The present study is concerned with developmental changes in the meaning of minority status to Jewish children. It attempts to diagnose self-other attitudes which may be related to group-belonging; to investigate the kinds of feelings of security, threat or anxiety, and the kinds of defenses which have developed; and to study the children's social ideologies.

SUBJECTS

One hundred fourteen children, from seven years to seventeen years of age, were studied. The distribution of the sample is given below:

	7–8 years	10–11 years	13–14 years	16–17 years
Girls	19	18	7	13
Boys	3	18	21	15
Total	22	36	28	28

All the children were attending one of two Jewish Community Centers located in Greater Boston in an area in which a high proportion of the population is Jewish. The children in the study come from homes of the low-middle and middle income levels. Most of the fathers are either white-collar workers or proprietors of small shops and businesses. The national backgrounds of the families are mainly eastern European. Twenty per cent of the parents are foreign-born. Orthodox religious practices are strictly observed in a few homes, but the majority of the parents are not very religiously observant.

PROCEDURE

Data were obtained through the use of a picture test and a questionnaire administered in group situations. (The youngest children were given help in writing their answers.) The picture test was given twice with an interval of three weeks between the two testings; the questionnaire was given at the second testing.

Measuring instruments.—In the picture test the subject is asked to judge children's 'character' from their photographs, and to choose children from the photographs whom he would like as friends and whom he would reject.

Each subject is given a 12 by 12 inch card, on which are photographs of eight boys and eight girls. Descriptions of behavior and personality are read to the group. The subjects are instructed to select the child who seems best to fit the description. After the character judgments are made, the subject is asked which child he would (would not) like (a) to have in his club, (b) to invite to a party, (c) to have as a work partner in school. On the first administration of the test the pictures are not identified in any way. On the second administration half of the pictures are identified as being Jewish children (left half of card) and the other half as being Christian (right half of card).

The test pictures were those which a group of adults had judged with equal frequency as 'Jewish' and 'not Jewish.' The pictures had also gone through preliminary testing on forty children who matched the pictures with character descriptions. Only pictures which were not over-chosen for any one description were used in the final test. Pictures later labeled as Jewish or Christian were equated as far as possible in terms of the pretesting.

The descriptions of behavior and personality characteristics were selected from studies of adult stereotypes of Jews. The test consisted of the following statements: (The stereotype which each describes appears in parentheses.)

1) This child tries to boss everyone else in the playground, and always has to have its own way. (domineering)

2) This child is very rich; the parents have just made a lot of money and try to show how rich they are. (nouveau riche)

3) This child is afraid to fight back when "picked on" by other children. (cowardly)

4) This child is sneaky. (sly)

5) This child is a show-off. (ostentatious)

6) This child is very smart, always gets the best marks in school. (intelligent)

7) This child doesn't speak English very well. The parents don't talk English at home. (foreign)

8) This child is "stuck-up." (conceited)

9) This child is never willing to share any of its things with other children. (selfish, greedy)

10) This child is not good in sports or games, is very clumsy. (lacking in physical ability).

11) This child always cheats when playing with other children. (dishonest)

12) This child studies all the time because the parents want it to get good marks in school. (intellectually over-ambitious)

13) This child doesn't mix very well; doesn't have any friends. (introverted)

14) This child is always talking about how much money it has and how much everything costs. (mercenary)

15) This child never waits its turn in line but always tries to push everybody around and get ahead of them. (aggressive)

16) This child is always loud and noisy. (loud and boisterous)

17) This child always sticks with its own group and doesn't like to play with children of other religions. (cliquish)

18) This child always talks with its hands. (uses gestures)

19) This child is always quarreling or starting fights. (argumentative)

20) This child always likes to know everybody's business, always "butts in" everybody's business. (prying)

At the second administration the following instructions were given:

> Several weeks ago you saw some pictures of boys and girls. Today we're going to see them again. Let's see if you can guess what these children are like just by seeing their pictures. Some are Jewish children; some are Christian children. The pictures on this side are Jewish, and on this side Christian.

On the first administration the test was scored by counting the number of times that pictures which were to be labeled Jewish on the re-test were chosen as fitting the behavior described. Assigning all descriptions to pictures which were later identified as Jewish would yield a score of 19; and assigning all descriptions to pictures later identified as Christian would yield a score of zero. (Item six was omitted from the score because it was the

11.8.70
J 101

only positive stereotype presented.) On the re-test the score was the number of times the pictures, now identified as Jewish, were chosen.

A questionnaire followed the picture test in the second session. The questionnaire was designed to elicit reactions concerning identification with the Jewish group. Situations pertaining to activities in school, in a club, in a recreation center, in the neighborhood, and in the community were described. The subject is asked to choose between Jewish and non-Jewish alternatives and to give reasons for his choice. The questionnaire follows:

Questionnaire

(Check (x) the answer you want. Then tell why you chose that answer.)

1. Suppose that there is going to be a new club here that you can join. What kind of name would you like it to have—a name that is Jewish or a name that is not Jewish?
   Jewish__________ Not Jewish__________
   Why?
2. If you could choose a badge or pin to wear to stand for the club, which would you like—a pin or badge that stands for something Jewish or for something not Jewish?
   Jewish__________ Not Jewish__________
   Why?
3. Suppose this new club is going to act out a story. What kind of story would you like to act out—a Jewish story or a story that is not about Jews?
   Jewish__________ Not Jewish__________
   Why?
4. If the music leader brought some songs for your club to learn, which would you choose—a Jewish song or a song in another language?
   Jewish__________ Not Jewish__________
   Why?
5. Suppose the new club can choose anything it wants to learn about. Would you choose something about Jewish life or about something that is not Jewish?
   Jewish__________ Not Jewish__________
   Why?
6. If you had to give a speech in front of your whole class in school what would you like to talk about—a Jewish subject or not a Jewish subject?

Jewish__________ Not Jewish__________
Why?

7. If you had $5.00 to give away to help poor people who needed money to buy food and clothes, would you give it to help Jews, or to help all poor people?
Jews__________ All poor people__________
Why?

8. Suppose you moved to a different town named Rockville, which would you like to join—the Rockville Neighborhood House or the Rockville Jewish Neighborhood House?
Rockville Neighborhood House__________ Rockville Jewish Neighborhood House__________
Why?

9. If you moved to a different street would you like only Jews to live on the street, or both Jews and non-Jews?
Jews__________ Jews and non-Jews__________
Why?

The situations on the questionnaire explore group identification in a variety of contexts. Only in a limited sense can the nine situations be thought of additively. Questions 3, 4 and 5 involve identification with religious or cultural aspects of the group. Choices on questions 1 and 2, on the other hand, are expressions of a more general desire to identify with, and to be identified as belonging to the Jewish group. Choice of a 'Jewish topic' for a speech in public school (question 6) requires an in-group choice in a situation in which, generally speaking, other content would be more relevant, and in which the child may anticipate some feelings of hostility toward the minority group. Choice of Jewish for a street on which to live or a Community Center to join (questions 8 and 9) involves a wish to restrict one's associations to Jews, a wish which undoubtedly reflects how the child perceives non-Jews and how he wishes to relate himself to society. Item 7 calls for the degree of sympathy felt for those who need help. Here many conflicts could be created concerning humanitarian feelings, feelings of special needs of the minority, resentments toward outgroups, reluctance to help those who may in turn be anti-Semitic.

An analysis of the data obtained from the picture test and the questionnaire follows.

FINDINGS ON THE PICTURE TEST

Since there was no identification of pictures as Jewish or Christian on the first testing, pictures from the two sides of the card should have been chosen with equal frequency (assuming that the two groups of pictures were equated, and except for possible position preferences on the page of photographs). The mean scores should approximate 9.5. On the second administration, to the extent that stereotypes of Jews are accepted, there should be a change in the direction of higher scores. The results are given in Table I.

None of the differences between the two testings approaches statistical significance; that is, there is no evidence, on this level of response, of acceptance by the Jewish children of common stereotypes about Jews, or biases toward Jews or Christians in terms of 'character' differences.

TABLE I.—MEAN SCORES ON CHARACTER JUDGMENTS ON PICTURE TEST

Age group	Test I	Test II	"t"	Significance level
7–8	10.00	9.59	.16	n.s.
10–11	9.78	8.95	.35	n.s.
13–14	9.34	9.62	.06	n.s.
16–17	9.63	9.70	.11	n.s.

It seems likely that the meaning of the test varied with the age of the children; that the older children, at least, were aware of the purpose of the test. It is possible, therefore, that their apparent lack of bias may stem from a deliberate attempt to appear unbiased. Such an attempt might be made by the child's alternating his choices on successive questions, assigning one stereotype to a Christian child, the next to a Jewish child, and so on. This possibility was tested by examining the correlation of successive items on the second testing. Eighteen correlations are possible between successive pairs of the twenty items. On these, nine were negative and nine were positive or zero for the seven- to eleven-year-olds' responses; whereas fifteen were negative and only three positive or zero for the thirteen- to seventeen-year-olds' responses. This preponderance of negative correlations for the older children, i.e., of instances in which group choice is alternated, suggests that the older children were motivated to answer impartially, to demonstrate their lack of prejudice.

The sociometric part of the picture test (choosing three and rejecting three children for potential club mates, for guests at a party and for co-workers) yields results similar to those reported above. Each choice of a picture of a Jewish child was scored one, and each rejection of a picture of a Christian child was scored one. The possible range of scores on choices and on rejections is from 0 to 9. The mean scores for each age group (Table II) indicate that about an equal number of Jewish and Christian children were chosen and rejected. No significant shifts occurred from the first to the second testing.

TABLE II.—MEAN SCORES ON SOCIOMETRIC QUESTIONS ON PICTURE TEST
(Second Test)

Age group	Number of Jewish children chosen	Number of Jewish children rejected
7–8	4.46	4.09
10–11	4.78	4.22
13–14	4.07	4.59
16–17	4.89	3.85

Summarizing the results from both parts of the picture test, the data show that these Jewish children have not accepted, or are unwilling to express stereotyped conceptions of Jews or Christians; they do not express a desire for friends in either Jewish or Christian groups exclusively, and they do not reject members of either group disproportionately.

These results should be interpreted in the light of two other kinds of data available in the study—the sociological background of the children and the testing situation, and the responses of the children to the 'why' questions on the questionnaire. The immediate neighborhoods are predominantly Jewish and the Community Centers in which the children were tested are wholly Jewish. The preferences expressed on the sociometric questions are, therefore, at variance with the actual circumstances of the children. The judgments of character on the picture test, as well as the sociometric responses, reflect what seems to be either the ideology which the subjects believe to be socially acceptable or the wished for state-of-affairs. The comparison of these responses and the 'why' responses is discussed later.

FINDINGS ON THE QUESTIONNAIRE

Responses to hypothetical situations described in the questionnaire bring out, to some degree, the children's values and philosophies, as well as their needs and conflicts related to minority group-belonging. Their choices of categories are analyzed first, followed by an analysis of the kinds of perceptions and motivations prompting their choices.

Choices of categories.—Although the proportions of choices in each category vary from situation to situation, there is one outstanding and consistent trend (Table III). There is a decided decrease in the per cent of children choosing the 'Jewish' alternative (the only exception being on the question of charity). The largest decrease occurs between the seven- and eight-year-old group and the ten- and eleven-year-old group. These differences are statistically significant in all instances except on the questions of charity and community center. The decrease in choice of Jewish alternative continues to the oldest group tested.

There is another contrast between the seven- and eight-year-olds and the older children. Whereas the seven- and eight-year-old children respond in a similar way to each of the questions from 1 through 6 (which call for overt identification with the Jewish group and choices of cultural content), the older children show considerably more variation from one situation to the next. Their choices of the Jewish alternative on these questions vary from eleven per cent to seventy per cent.

The steepest drop with age in choice of Jewish occurs on the question of a speech in school. The ten- and eleven-year-olds and the seven- and eight-year-olds attend the same schools (in which there is a high proportion of Jewish children and in which there are both Christian and Jewish teachers); however, only slightly more than one-third of the ten- and eleven-year-olds

TABLE III.—CHOICE OF ALTERNATIVES ON QUESTIONNAIRE IN CATEGORIES—JEWISH, NOT JEWISH, BOTH GROUPS
(Per Cent of Children)

Question	Age Group	Jewish	Not Jewish	Both Groups	P*
1. Name for new club	7–8	95	5	0	
	10–11	59	30	11	.01
	13–14	36	43	21	.05
	16–17	43	43	14	n.s.
2. Pin or badge for club	7–8	100	0	0	
	10–11	70	24	6	.01
	13–14	43	46	11	.05
	16–17	57	21	22	n.s.
3. Story to dramatize in club	7–8	91	9	0	
	10–11	49	40	11	.01
	13–14	44	41	15	n.s.
	16–17	32	54	14	n.s.
4. New songs to learn in club	7–8	82	18	0	
	10–11	43	43	14	.01
	13–14	25	64	11	n.s.
	16–17	36	53	11	n.s.
5. Something to study in club	7–8	95	5	0	
	10–11	59	27	14	.01
	13–14	46	29	25	n.s.
	16–17	57	39	4	n.s.
6. Topic for speech at school	7–8	95	5	0	
	10–11	39	50	11	.01
	13–14	14	79	7	.05
	16–17	11	82	7	n.s.
7. Money to charity	7–8	27	—	73	
	10–11	32	—	68	n.s.
	13–14	17	—	83	n.s.
	16–17	70	—	30	.01
8. Joining Center in new town**	7–8	77	—	23	
	10–11	54	—	46	.10
	13–14	48	—	52	n.s.
	16–17	44	—	56	n.s.
9. Moving to new street	7–8	59	—	41	
	10–11	25	—	75	.01
	13–14	21	—	79	n.s.
	16–17	11	—	89	n.s.

* Chi square differences were computed between successive age levels, dichotomizing the data—Jewish, and not Jewish or both groups. Chi squares of 6.635 are significant at .01 level; of 5.412 at .02 level; of 3.841 at .05 level.

** Downward trend is gradual so that adjoining age groups do not differ significantly. Chi squares between 7–8 and 13–14, and 7–8 and 16–17 are significant at .05 level.

express a readiness to discuss a Jewish topic before the class compared with ninety-five per cent of the seven- and eight-year-olds who choose to do so. Children from thirteen to seventeen years attend schools which have many more non-Jewish children. At these age levels few select a Jewish topic (fourteen per cent and eleven per cent of thirteen- to fourteen-year-olds and sixteen- to seventeen-year-olds, respectively).

In joining a community center in another town, and in moving to a new street, choices of Jewish are not so frequent as in the preceding questions, even among the youngest children. Again increasing age brings fewer ingroup choices. The trend is from seventy-seven per cent of the youngest children to forty-four per cent of the oldest group choosing a Jewish community center, and from fifty-nine per cent to eleven per cent choosing a new street with only Jewish residents. These responses, like those on the picture test, indicate a desire for associations which are not wholly with the ingroup.

The question concerning money to be given to charity evokes a pattern of responses which is very different from that of the other questions. The majority of children below the sixteen to seventeen age level choose to give to all poor people rather than to Jews alone. With the sixteen to seventeen year olds the trend is abruptly reversed. Seventy per cent of this group choose to give their charity solely to Jews as compared with seventeen to thirty-two per cent of the three younger groups who express this preference. The explanation for this shift is to be found in the motives involved in the choice. To many of the younger children social charity is a circumstance which calls for 'fairness.' Their loyalties to ingroup are tempered by stronger needs to be fair and equal to all. Whereas, in the other situations, the choice of Jewish does not imply an attitude of selfishness, a similar choice in the charity question does. Perhaps fairness is no less a part of the older children's ideology; however, choice of Jewish on the charity question is made on the basis of some special factors which overshadow fairness considerations. It is not unlikely that the older children are more keenly aware than their younger companions of the plight of Jewish refugees and feel it more imperative to contribute to their relief. Also, the children of the oldest group are at the age when they are beginning to think seriously of their future careers, of getting a job or of going to college. These considerations, too, may heighten the feeling of being imbedded in a hostile non-Jewish environment, and strengthen the need to retaliate or protect the ingroup. These interpretations are supported in the responses to the 'why' questions, discussed below.

Reasons given for choices.—In analyzing the reactions of these children to ethnic group membership, it is important to consider to what extent their responses are 'normal' reactions to group identification with 'normal' developmental changes, and to what extent their responses are conditioned by minority status and represent rejections and fears concerning their own group. With increasing age, the child's interests and identifications are expected to increase and to extend to a variety of primary and secondary groups. His family and its cultural group constitute one identification and interest among many others. On this basis, relatively fewer ingroup choices should be expected as the child gets older, regardless of the minority status of his family group. Ideally, a part of growing up in a democratic society would seem to involve both the freedom to develop ingroup loyalties, and the freedom to seek congenial interactions in a culturally diverse society. On the other hand, the special pressures to which a minority group is subjected may give rise to other needs with respect to group belonging.

Four general types of perceptions and motivations appear in the reasons which the subjects give for their choices of alternatives on the questionnaire: (1) Judgments of groups as 'better than' or 'not as good as' are expressed. In these statements there is often a kind of absolutism, "I'm Jewish, therefore, I like Jews best." (2) Specific satisfactions or dissatisfactions derived from group membership are given as reasons for choice. (3) A desire for cultural content and associations including more than Jewish group, or, on the contrary, excluding Jewish culture explains other choices. A philosophy of group relations is sometimes expressed. (4) Apprehension concerning Jewish and non-Jewish interactions appears in many responses. Here awareness of minority status or group conflict is explicit, and feelings of tension and insecurity are evident ("I don't think we would mix well," "There's too much anti-Semitism and that would only arouse it," "I would not be prejudiced in favor of my own religion and would be as tolerant as possible.")

These answers to 'why' show as marked developmental trends as the choices of alternatives. The seven- and eight-year-olds most frequently give the kind of reason described in (1) above. The thinking may be paraphrased as follows: "I am Jewish; therefore, the things I do and like should be Jewish." It is a categorical 'like' or 'don't like,' a 'good' or 'bad' with respect to either group. This same perception tends to persist throughout the child's answers; specific factors in the situations seem to have little influence. It can hardly be said that an ideology is the basis for these reactions. They seem rather to be the result of limited differentiation in the child's social world, and of an absolutist kind of reasoning, not uncommon in young children.

The seven- and eight-year-olds' rule of "I like Jewish because I am Jewish" is broken at two points in their questionnaires. Simple ideologies appear on the charity question and on the questions of a community center and a street. On the charity question (in which seventy-three per cent of the children decide to give their money to help all poor people) the ideological support is one of 'fairness,' ("I would like to give to all the people because it's not fair to give to one."). As noted earlier, at this age level, dividing equally constitutes fairness, and other contingent factors (the possible anti-Semitism of the people being helped, the need of Jewish refugees) are either unknown or they are not seen as related to fairness and are not considered in arriving at a decision.

The other ideology which appears among the seven- and eight-year-olds is that all children should play together. On questions of a community center and a street of residence, twenty-three per cent and forty-one per cent, respectively, choose Jews and non-Jews with this ideology verbalized.

The seven- and eight-year-olds never speak of prejudice against their own group and rarely (four per cent) indicate apprehension about identifying with their group or about associating with persons who do not belong to their group.

As noted above, at this age level the varying requirements and implications of the several situations have relatively little effect in altering the child's basic response pattern. This tendency is illustrated in the seven-year-old boy's responses which follow: (Club name—Jewish) "I like Jewish because I am a Jewish boy." (Club pin—Jewish) "I like Jewish because I am a Jewish boy." (Club story—Jewish) "I would like to be in a Jewish act because I am Jewish." (Club songs—Jewish) "I would like a Jewish song because I am Jewish." (Something to learn about—Jewish) "I would like to learn about Jewish life because I am a Jew." (Speech in class—Jewish) "I would like a Jewish subject because I am a Jew." (Money to poor—Both groups) "I would like to give to all the people because it's not fair to give to one." (Neighborhood House—Jewish) "I would like to live in a Jewish neighborhood because I am a Jew." (Street—Jewish) "I am a Jewish boy and like Jewish people."

If it can be assumed that, earlier in their development, the older children in the sample were substantially similar to these seven- and eight-year-olds, then the differences in responses from one age level to the next can be interpreted as the resultants of developmental factors and social experiences.

One marked change with increasing age is in the absolutist responses of "I like because I am." These responses decrease and all but disappear. Much greater variety of point of view and motivation appears. Older children, who have perhaps as strong an ingroup orientation as the seven-year-old quoted, (or an equally uncompromising outgroup orientation), have learned to support or rationalize their position in varied ways, depending partly upon the stimulus situation. Thus, a teen-ager who always chooses the Jewish alternative does so because "the Jewish language is nice," "you know more Jewish people," "it's good to help your side," "you should practice your religion," "if non-Jews are around there might be a fight." He has maintained the same position as the younger child but with 'good' reasons. The same process could be illustrated with an outgroup orientation.

The following motives appear most commonly in the older children's reactions: Feeling a duty or desire to learn about their culture or to practice its customs ("If you are Jewish you would like to learn your history") motivates many of the choices of Jewish content. Thirty to forty per cent of the responses of the ten- and eleven-year-old children and fifteen per cent to thirty per cent of the responses of the two older groups include this motive in the questions concerning club content.

Gaining a feeling of solidarity or belonging by being with ingroup members explains the preference of about twenty per cent of the older children for a Jewish center rather than a community center. In answers to all of the questions, the strivings for security and the feelings of anxiety become increasingly evident with age. Anxiety regarding the reactions of non-Jews is sometimes expressed in a child's statement that a Jewish name or badge is a handicap, or that the choice of Jewish may be viewed as prejudice, or that non-Jews would feel out of place in a club with a Jewish name or Jewish content. Note the sudden and sharp rise in this response from the youngest age level. Only four per cent of the seven- and eight-year-olds indicate this anxiety. At the ten- and eleven-year level the percentage has risen to thirty-eight per cent; at the thirteen- and fourteen-year level to forty-three per cent and at the sixteen- and seventeen-year level to fifty-nine per cent. Several illustrations are given below:

"(I chose Jewish) to show that I'm not afraid of any race prejudice, and perhaps to teach children of other races something about mine." "Because if it has a Jewish name some other clubs may not want activities with us, but if it is not Jewish we can socialize better." "Because of my religious persecution I wouldn't dare speak on Jewish matters, then again it would probably start awful arguments."

Just as there is increasing sensitivity with age, there is an increasing desire for wider cultural content and associations. The following motivations are typical on questions of the community center and the street, and the content of the clubs: "They should have a play about Jews and then learn songs that are not Jewish." "(On a street of Jews and non-Jews) you could all have fun together."

The occurrence of these two kinds of responses within the same children expresses the conflict which confronts the minority child, conflict which is often complicated by other factors as well.

By evaluating the 'sum' of the responses of each child, it appears that in a number of the children explicit ideologies regarding intergroup relationships have developed. With few exceptions this ideal is one of 'tolerance,' of harboring no prejudices. The following statements are representative: "I like all and hold no distinction between one race and another." "I have no race discrimination in my heart." "It's better to get to know all kinds of people because you're less likely to be prejudiced." "If we only picked Jewish this wouldn't be a very good world and surely not a democratic world."

The less conciliatory point of view appears only in a few cases; thus: "I'd meet Jewish boys, not that I despise Christians, just that I prefer Jews."

It was pointed out earlier that the several situations of choice on the questionnaire were evaluated more individually by the older than by the younger children. For the youngest children it appeared to be a matter of choosing Jewish or not Jewish, with the characteristics of the specific situations being of little importance. For the older subjects the nature of the situation more often entered into the choice: There are 'appropriate' situations in which a Jewish child should (or would) select Jewish cultural content and Jewish associations, ("If there is all Jewish kids there is no sense in having a not Jewish badge"); however, there are other circumstances in which an ingroup distinction is irrelevant ("Because we don't study Jewish in school," "When it comes to that (choice of a street) I'd rather have religion out of it.") It is at the point of such differentiations that one is able to begin to discern basic differences in the social orientations which these children are developing as consequences of minority membership. The majority of the children are neither wholly ethnocentric nor wholly rejecting of their ingroup. There are a small number of the older children, however, for whom each of the situations constitutes a stimulus for having, or for seeking, a reason which justifies either a rigid rejection of all that is ingroup or a rigid adherence to all ingroup choices. This study does not reveal the kinds of conditions which give rise to these extreme reactions. Nor is it possible to mark their beginning in terms of age. Perhaps some of the children, now sixteen or seventeen, have persisted in the kind of reactions characteristic of the seven-year-old; perhaps others have arrived at their extreme position from less extreme precedents, through the effects of certain environmental or personality factors. This problem requires further research.

In emphasizing the kinds of philosophies present, one may be misled into thinking that each of the children has worked out a consistent point of view which gives him 'the answers' in any situation of decision. Such is not the case. The child's uncertainties and the conflict mentioned earlier must be examined more closely. The conflicts which these children show are not simply between desires for free social interaction and fear of rejection. The conflicts broaden to include over-anxiety about being considered ethnocentric for showing any interest in their own group. Some children have turned completely away from any ingroup loyalities, describing all that is not Jewish as 'better,' 'more interesting,' and the like. A conflict which appears particularly among the sixteen- and seventeen-year-olds (on one or more questions in twenty-one per cent of the group) is implied in the notion that being Jewish precludes being an American: "People who didn't know the club would think that we are self-conscious of being Jews and not Americans," "An American name would be preferred," "I am living in America and my country must come first." To add to the dimensions of conflict in the two older age groups, stereotypes about Jews appear for the first time: "Because sometimes Jewish people are awfully noisy and anyway it's better to get along together than to stick to your own little group," "Non-Jewish peoples' streets look cleaner," "I trust more Jewish people than Christian," "Jews do gossip a lot," "I get along with both just as well and while a Jewish neighborhood is noisy and alive a non-Jewish neighborhood is quiet and restful." More stereotypes have been expressed by the older children in the projective situations of the questionnaire than appeared in their more self-conscious responses to the picture test. This difference in itself perhaps reflects ambivalences of believing and not believing the stereotypes.

The sixteen- and seventeen-year-olds, it will be recalled, reacted differently from the younger groups in responses on charity. The seventy per cent who choose to give their charity to Jews do so with their primary reasons being: "I would help my own first," and "Jewish people in Europe are especially in need. It is our duty to help them." In these responses one can discern the factors (discussed earlier) of awareness of needs of Jewish refugees and of defensiveness against a hostile non-Jewish world. In light of the other responses by the older children—responses filled with conflict—this act of defense is readily understood.

DISCUSSION OF FINDINGS

This documentation of Jewish children's reactions to group membership furnishes some indication of the kinds of effects which minority membership has on children. From the questionnaire responses one glimpses some of the significant perceptions and motivations of these children in their attempts to adjust to the ingroup and to the demands of society.

It is quite probable that some of the developmental changes observed reflect increased social maturity and are not peculiar to the minority situation, particularly such changes as the following: Undifferentiated perceptions and lack of relativity of judgments to meet changing requirements of different social situations characterize the responses of the youngest children, but give way to increasingly differentiated perceptions, more varied goals and greater relativity of judgments in the older children. This finding is consistent with developmental data on problem-solving and concept formation. The trend away from choosing exclusively within the cultural group of the family is probably also 'normal' growing up. To what extent this process is to be expected in children belonging to any group, and at what point it represents anxiety and rejection of the ingroup because of social pressures, are crucial questions in diagnosing the effects of minority membership upon children.

The questionnaire responses leave no doubt but that early in childhood (appearing among the seven- and eight-year-olds in this study, but assuming more marked proportions by ten years) minority status begins to be felt. Increasingly with age the children begin to see their relations to their ingroup and to non-Jews in both ideological and 'practical' terms. At the same time that their contact with non-Jews becomes more extensive and their absorption in the prevailing majority culture more pronounced, being a member of a minority group becomes increasingly a source of distress. The social restraints on normal strivings toward integration into the general culture of the school and community result in the well-known consequences of frustration. Judging from the children's responses, fear and uncertainty with resulting anxiety in social relations are most frequent outcomes. The choice of defenses, as it were, has not yet been made by many children. A few of the teen-age children have turned to the protective philosophy of "my own group is best and first, and in it I seek all my satisfactions." A few have taken the opposite course, "my own group is inferior; to choose it or to recognize worth in it is contrary to democracy."

Some of the children have consciously thought-out ideologies with respect to group relations, in most instances supporting democratic acceptance of differences and rejecting prejudice. It would be interesting to compare minority children with non-minority children in this respect. Is there greater social consciousness, greater striving for democratic group relations on the part of the minority children? May one not expect to find, also, within the minority, retaliatory reactions in which there is overvaluation of ingroup loyalties and depreciation of outgroups and exaggerated lines of distinction between groups? It is well known that these reactions, as well as feelings of deep social consciousness, develop in some adult minority members. That this reaction occurs infrequently among the children studied may be due to the particular sampling or to the fact that the retaliatory reaction may develop more often later, after the individual's struggle for acceptance has met many defeats and when his central goals are perceived as inaccessible because of social discriminations against his group. Perhaps not many of the children have reached this stage.

The responses to the open-ended 'why' questions revealed many of the tensions which the minority situation holds for these children. Their responses to the character judgments of the picture test and their replies to sociometric choices, which appear so 'well adjusted,' so free of tension, are very significant data to be compared with their responses to 'why.' They are significant for several reasons: (a) because they show the children's public reactions on matters of group-relations, their official, or perhaps ideal, ideology; and (b) because such reactions are often made the basis for diagnosing the situations of school and community life in which varied ethnic groups are involved; where it is concluded that "all is well," "we have no problems of prejudice here; we all get along together." The error is obvious.

Some years ago child psychology, education and sociology alike became greatly concerned about adolescence as the period of 'storm and stress' in our culture. Thereupon, its problems and remedies were studied in research and practice, with resulting better understanding and treatment. Is there not equal cause for concern about the social-psychological position of 'storm and stress' of children of minority groups in America?

More comparative research is needed on other minorities in our culture. Most urgently indicated is research on methods of modifying these social mores which are contrary to democratic ideals and which preclude equality of opportunities for social development of all children.

A STUDY OF SOCIAL SENSITIVITY (SYMPATHY) AMONG ADOLESCENTS

WALTER LOBAN

School of Education
University of California

The initial problem of this research was to explore techniques designed to discriminate between adolescents who are highly sensitive to the feelings of other people and those adolescents whose sensitivity to the feelings of others is demonstrably low. Then, having located two such groups, the next step was to study these two unlike groups with reference to attitudes, needs, behaviors, or idiosyncrasies which might help to explain their differences in social sensitivity.

Arrangements for this research were concluded with seven *public schools, a private school, and a reform school. These schools were chosen with the purpose of obtaining a sample representing the broader universe of American adolescents with* reference to sex; rural-urban distribution; socio-economic condition; race; as many of the representative religious faiths of America as possible; and a majority fairly characteristic of the so-called average American boy and girl, all in reasonably typical proportions. The sampling consisted of two hundred thirty boys and two hundred girls of whom eighty-eight per cent were Caucasian; eleven per cent, Negro; and one per cent, Oriental. The rural-urban and socio-economic distribution is given in Table I where students are arranged in seven socio-economic levels according to their parents' or guardians' occupations as rated by the Minnesota Occupational Scale (*1*). Comparisons are also given with two earlier studies.

The nine schools coöperating in this study met these conditions remarkably well. All together, four hundred thirty adolescents ranging from Grade Eight through Grade Twelve participated in the study, three hundred seventy-six of them representing public school students and fifty-four of them representing private and reform schools.

Two measures constituted the basis on which the subject's social sensitivity was determined. The first of these, a sociometric instrument, contained descriptions of behavior low or high

TABLE I.—MINNESOTA OCCUPATIONAL SCALE: PER CENT OF ADOLESCENTS IN SEVEN SOCIO-ECONOMIC LEVELS (COMPARED WITH TWO EARLIER STUDIES)

Socio-economic level	1947–8 % This Study	1940 % Hansen (*2*)	1920 % Goodenough and Anderson (*3*)
I) Professional	3	3	3
II) Semi-professional and retail business	9	7	5
III) Clerical, skilled trades and retail business	17	14	14
IV) Farmers	13	15	19
V) Semi-skilled occupations, etc.	36	24	27
VI) Slightly-skilled trades, etc.	18	15	13
VII) Day-laborers	2*	22	19

* Although 1947–8 was a fairly prosperous economic period, it is reasonably certain that this per cent should be higher; there is evidence that some pupils in this category enhanced their parents' jobs, placing the jobs in V and VI.

in sensitivity. For example, one item read: "X is a person who is unusually sympathetic and sensitive to how other people feel. X would never knowingly hurt the feelings of another person. Who is X in your class?" The subjects filled in the names of classmates who exemplified the behavior of each item. Some items were dummy items which concealed the purpose of the instrument. The instrument was titled *Is It Anybody You Know?*[1]

Is It Anybody You Know? was scored by tabulating the number of times each student received mention for (1) items which described sensitive, sympathetic personalities and (2) items which described insensitive, inconsiderate personalities. Sensitive items became plus scores; insensitive items were given minus signs. Subjects whose neutral personalities gave them little or no mention gravitated toward zero. Since class populations differed, all scores were divided by class size in order to make them comparable. On this distribution the subjects varied from a −73 to a +94.

---

[1] Copies of this instrument are available from the writer.

A second measure was a simple rating scale on which their teachers arranged the subjects on seven steps of social sensitivity ranging from 'exceptionally sympathetic and considerate' to 'ruthless, cruel, brutal.' This rating by teachers took place at the close of the school year during which the teachers, aware of the purpose of this research, had learned with a sharper attention than usual to observe the social behavior of their pupils for manifestations of sensitivity. Many of these teachers, also, were the advisors, homeroom directors, or counselors of the subjects.

As a result of these two measures, scores were available which recorded (1) the judgments of pupils concerning each other's sympathetic behavior and (2) the judgments of teachers on the same matter. Both judgments were valuable. The pupils knew each other better and saw more of one another's behavior outside of class and school. On the other hand, the teachers had the advantage of maturity and experience. Furthermore, the teachers had been alert to the whole problem of sympathetic behavior ever since they had begun to work on this investigation at the opening of the school year.

The judgments of pupils and teachers were given equal weight. A composite distribution of teachers' and pupils' judgments was made for the total number of cases; from this sixty cases were selected at each extreme in social sensitivity.[2] The difference between the two groups on the combined pupil-teacher measures was statistically significant at the one per cent level.

A third measure, the *Hawthorne Group Test of Cruelty-Compassion* (*4*), was a standardized test used as a check on the other two instruments. Two forms of this test were used. The form for boys was identical with Hawthorne's standardized test. For girls this form had to be adjusted in order to exclude choices that were solely masculine in nature. The same number and the same kind of items appear on both forms. When the t-test is applied to the difference in means between the most sensitive boys and the least sensitive boys, as measured by the combined pupil-teacher rating, the difference proves significant at the one per cent level of confidence. The same test for significance also achieved the one per cent level for the difference between sensitive and low sensitive girls.

---

[2] The distribution was approximately normal, with sixty-eight per cent falling within one standard deviation on either side of the mean.

Throughout the remainder of this investigation the two groups representing the extremes of social sensitivity among the subjects were studied separately from the large middle group of adolescents. However, all subjects were further studied by a long series of measures falling mainly into two categories:

1) Measures designed to learn as much as possible about the subjects' responses to ten selections of literature in which the authors' intention is clearly to evoke sympathy.[3] These 'measures,' constructed for this research, were: content analysis of free responses, questionnaires, and social-distance scales applied to the main characters in the stories.

2) Measures designed to learn as much as possible about the subjects themselves—their intelligence, emotional needs, socio-economic status, religious faith, church attendance, acceptance or rejection by other adolescents, home conditions, reading ability, health, mental stability, values and ideals, race, sex, and so forth. For as many of these measures as possible, a mathematical score was obtained for each individual.

Sociograms were used to determine the subjects' acceptance and rejection by their peer group (*5*). For socio-economic status, the Goodenough-Anderson *Minnesota Occupational Scale* (*6*) was applied to the jobs of the subjects' parents. Eight emotional needs were investigated by a form of the *Self-Portrait* developed by Louis Raths (*10*). Intelligence was measured by the *Otis Quick-Scoring Mental Ability Tests* (*9*), and reading by *Traxler Reading Test* (*11*). Questionnaires dealt with such factors as church attendance, family size, race, ideals, values, choices of books, and the three wishes a subject would choose. One hundred and sixty-two subjects, including twelve subjects chosen for special case studies, took the *Minnesota Multiphasic Personality Inventory* (7). A fictitious list of book titles as well as several tests of choices concluded the testing program by probing still further at the preferences of high sensitive and low sensitive persons.

---

[3] The teachers read aloud to their classes the following ten stories: (1) *A Mother in Mannville* by Marjorie Kinnan Rawlings; (2) *Yours Lovingly* by Eugenie Courtright; (3) *A Start in Life* by Ruth Suckow; (4) *The Kiskis* by May Vontver; (5) *The Beginning of Wisdom* by Rachel Field; (6) *Miss Brill* by Katherine Mansfield; (7) *The Horse* by Marian Hurd McNeely; (8) *The New Kid* by Murray Heyert; (9) *Prelude* by Albert Halper; (10) *That's What Happened to Me* by Michael Fessier.

Many precautions were taken to prevent unwarranted errors. Specialists and experts were consulted on such matters as measurement, sociometrics, and inter-group relations. Preliminary tryouts of all the measures were carried out with adolescents in the Minneapolis Public Schools, and numerous revisions were made on the basis of those trials. Teachers in the experiment were visited and briefed in advance, and students were repeatedly assured that their responses would in no way affect their status in school. Particular care was taken to treat all the adolescent subjects with the same deference and respect that any cultured adult shows toward another adult.

To test the significance of the difference between the various means of the two groups of adolescents, the t-test was used. The number of results which proved to be significant at the one per cent level indicates that the two groups of adolescents were definitely different from each other. The most reasonable conclusion is that the basis on which they were grouped—sensitivity to others—was a valid basis.

In this study the matter of 'spread' or 'variance' became highly important because of the presence of both boys and girls in the two groups. As soon as the data were tabulated, it became obvious that more girls than boys appeared in the group of highly sensitive adolescents.[4] The opposite was true of the low sensitive group, where boys predominated.[5] Since these sex differences were significant at the one per cent level, it became necessary in all procedures to make certain whether or not boys and girls in the same classification could be treated as a single group. To pool boys and girls, it was necessary to show not only that they were drawn from larger groups whose means were the same but also that they came from equally variable populations in regard to the measure being investigated.

---

[4] There were thirty-six girls and twenty-four boys in the high-sensitive group.

[5] There were forty-four boys and sixteen girls in the low-sensitive group.

For example, the girls in the low sensitive group spread over a wide range of social and economic prestige while the inconsiderate boys clustered closely around the mean. Because these differences in variability between the inconsiderate boys and girls were significant, it became impossible to compare these boys and girls as if they were a single group capable of being compared with the sensitive adolescents on this matter of socio-economic status. Instead, it was necessary to separate them on this matter of socio-economic status. Comparisons, of necessity, had to be made between inconsiderate girls and sensitive girls, between inconsiderate boys and considerate boys. Because the f-test showed a difference of variance in the one area of socio-economic status, it became necessary to test the variance on every measure to determine whether or not it was justifiable to pool boys and girls together.

Furthermore, for any study of social sensitivity, an adequate comprehension of the meaning of measurement is important. True measurement requires units that are constant, interchangeable, and additive. However, in most psychological and educational measurement, the zero point is not known. The most that can be said is that certain individuals vary in the same direction (or in opposite directions) and that we do not claim anything more than a rank ordering of our cases. In the measures and scales of this investigation, the scores have relative meaning only. They are useful scores in that they determine the individual's relative status in the group with which he is compared, but they are never to be considered as absolute scores nor as comparable in the manner of scores in a physical scientist's ideal of true measurement.

RESULTS

These measures made it possible to rank the adolescents along a continuum ranging from ruthless and inconsiderate behavior to sympathetic and thoughtful behavior of the highest order and to select two extreme groups of adolescents who varied greatly on this continuum. Some of the differences between these two groups are presented in Tables II and III.

1) As assessed by the Raths' *Self-Portrait,* an important emotional difference between very sensitive adolescents and low sensitive adolescents proves to be their feeling of anxiety con-

TABLE II.—DIFFERENCES AND SIGNIFICANCE OF DIFFERENCES BETWEEN GROUPS OF ADOLESCENTS VARYING IN SOCIAL SENSITIVITY AS MEASURED BY THE T-TEST

Variable	Group	Number	X	S.D.	S.E. diff.	t	Level of Significance
1. Hawthorne Group Test for the Measure of Cruelty-Compassion	High sensitive girls	35	122.6571	8.94			
	Low sensitive girls	16	113.5000	12.71	2.8833	3.18	one per cent
	High sensitive boys	24	98.6250	11.46			
	Low sensitive boys	40	86.0500	11.22	2.96676	4.239	one per cent
2. The Need to be Free from Anxiety	All high sensitive	60	−4.0333	3.90			
	All low sensitive	60	−0.9167	4.20	.7218	−4.318	one per cent
3. Intelligence	All high sensitive	60	107.6833	10.28			
	All low sensitive	58	103.9310	10.83	1.9608	1.914	not significant
4. Socio-economic Status	High sensitive boys	24	3.5833	1.29			
	Low sensitive boys	42	4.5952	1.15	.3133	3.230	one per cent
	High sensitive girls	36	4.2500	1.40			
	Low sensitive girls	16	3.8123	2.00	.4941	.885	not significant
5. Peer Group Acceptance	All high sensitive	60	14.3167	7.50			
	All low sensitive	60	−0.2167	11.42	1.8136	8.014	one per cent
6. Reading Ability	High sensitive boys	24	57.0000	31.66			
	Low sensitive boys	41	44.6341	30.24	8.0330	1.539	not significant
	High sensitive girls	35	64.6000	29.95			
	Low sensitive girls	16	66.3125	27.55	6.5820	.260	not significant
	All high sensitive	59	61.5085	30.89			
	All low sensitive	57	50.7193	30.07	5.8036	1.859	not significant
7. Ideal Self	For this variable, the Behrens-Fisher d-test was used. For the results, see item 4 in the summary statements.						

TABLE III.—DIFFERENCES AND SIGNIFICANCE OF DIFFERENCES BETWEEN GROUPS OF ADOLESCENTS VARYING IN SOCIAL SENSITIVITY AS MEASURED BY THE CHI-SQUARE TEST

Variable	Group	Number	X^2	Level of Significance
1. Sex	High sensitive boys	24		
	High sensitive girls	36		
	Low sensitive boys	44		
	Low sensitive girls	16	13.57	one per cent
2. Content Analysis of Response to Ten Short Stories	All high sensitive	60		
	All low sensitive	60	20.7108	one per cent
3. Church Attendance	All high sensitive	60		
	All low sensitive	59	2.341	not significant
4. Size of Family	All high sensitive	59		
	All low sensitive	60	5.576	not significant

cerning their behavior. As a group, the sensitive adolescents are more concerned over their relations with other people, and they worry more and longer over behavior which the least sensitive adolescents quickly forget. The high sensitive adolescent is much more aware of his own limitations, inadequacies, and failures in human relations. This difference was significant at the one per cent level.

2) Socio-economic status proved to be an important factor in influencing the social sensitivity of adolescent boys but not that of adolescent girls. Low socio-economic status is less conducive to sympathetic behavior for adolescent boys than average or good socio-economic conditions. It is probable that boys of low economic status, more directly faced with the problem of economic struggle, find it more necessary than girls to suppress tendencies toward sympathy and to adopt an attitude of 'every man for himself.' This difference for boys was significant at the one per cent level.

3) A larger number of adolescent girls proved to be highly

sensitive as compared with adolescent boys. This difference was significant at the one per cent level.

4) A significant difference between these adolescents is the greater inclination on the part of the least sensitive to approve persons "who run their own lives, who are reckless, independent, restless, free and impatient of all control and law . . . who are ungovernable, superior and powerful, free from the need of considering how others will react to what they do." For boys this difference is significant at the one per cent level; for girls it is significant at the five per cent level. This difference was assessed by means of a list of choices called *My Ideal Self*.

5) No significant difference in intelligence was found to exist between adolescents who were highly sensitive and those who were least sensitive.

6) Church attendance and denomination have no relation to the social sensitivity of adolescents, but the case studies reveal that the quality and intensity of religious experience do have a relation (whether this is the cause of sensitivity or the result is not certain).

7) The most sensitive adolescents are clearly more popular with their peers than are the least sensitive adolescents. This difference in popularity is also statistically significant at the one per cent level.

8) In every one of the literary measures used with all the adolescents, a persistent tendency to identify with literary characters most closely resembling one's self consistently appears.

9) No significant difference exists between the two groups in regard to reading ability as measured by the *Traxler Reading Tests*.

10) There is a persistent tendency for the highly sensitive adolescents to show a greater interest in books and choices that deal with idealistic, esthetic, and sympathetic themes. The least sensitive adolescents lead in an interest in books and choices that emphasize cruelty. Both groups tend to be about equally interested in material success and equally disinterested in a book that exalts prejudice and intolerance. Although the significance of these differences has not been demonstrated statistically, their consistent reappearance throughout several different tests is worthy of note.

11) On free responses, written directly after listening to each

of ten stories intended to evoke sympathy, the highly sensitive and low sensitive groups varied in the number of concepts which competent judges had declared a sympathetic reader with insight would notice. This difference, with the higher scores going to the most sensitive group, was significant at the one per cent level.

12) No differences in race developed in respect to social sensitivity in this study.

13) There is a tendency on the *Minnesota Multiphasic Personality Inventory* for highly sensitive adolescents to be more stable and to consider themselves in better health than the least sensitive adolescents.

14) No significant difference between the two groups was found in respect to size of family or position of child in order of birth.

15) On most of the measures in this study, girls vary more than boys. One possible explanation may be that the actual biological, constitutional basis on which personality is built in human beings may be no different in girls than in boys. However, exposed to greater social demands to become sympathetic, a group of girls may be less homogeneous than boys when both groups are selected according to the very factor which requires greater modifiability on the part of girls than of boys (*8*).

This research it is hoped will have value in the development of a study of human relations and in the education of tolerant, sensitive citizens. Underlying the problems of inter-racial and intercultural strife with which the public schools deal, lies the basic problem of social insight and sympathetic awareness of the feelings and thoughts of other minds. A technique of measuring this quality of social sensitivity should make it possible to take the next important step. By measuring groups before and after various kinds of sensitizing experiences, a more secure knowledge may be gained of what education is most effective in promoting better social sensitivity.

REFERENCES

1) Florence Goodenough and John E. Anderson, "Minnesota occupational classification scale" in *Experimental Child Psychology*. New York: The Century Company, 1931, pp. 234–238 and 501–12.

2) Hansen: quoted by Institute of Child Welfare, University of Minnesota, on a mimeographed sheet. No further reference given.

3) *Op. cit.*

4) Joseph Wyman Hawthorne, "A group test for the measurement of cruelty-compassion: A proposed means of recognizing potential criminality," *Journal of Social Psychology*, III, pp. 189–229.

5) Helen Hall Jennings, *Sociometry in Group Relations*. Washington: American Council on Education, 1948.

6) *Minnesota Occupational Classification Scale*. This scale and its derivation are described by Florence Goodenough and John E. Anderson in *Experimental Child Study*. New York: The Century Company, 1931, pp. 234–238 and 501–512.

7) *Minnesota Multiphasic Personality Inventory*, Starke R. Hathaway and J. Charnley McKinley, *Booklet and Manual*, Revised Edition. New York: The Psychological Corporation, 1943.

8) Lois Barclay Murphy, *Social Behavior and Child Personality*. New York: Columbia University Press, 1937. This research is exceptionally valuable for the insights into sympathy which it provides.

9) *Otis Quick-Scoring Mental Ability Tests* by Arthur S. Otis. Yonkers, N. Y.: World Book Company, Beta Test for Grades 4 through 9. Gamma Test for high schools and colleges.

10) Louis Raths, "A test of emotional needs," *Educational Research Bulletin*, XXXVI, No. 1, Jan. 15, 1947, pp. 14–17.

11) *Traxler Silent Reading Test Form 1* was used for Grades 8, 9, and 10. *Traxler High School Reading Test Form A* was used for Grades 11 and 12. Devised by Arthur E. Traxler. Bloomington, Illinois: Public School Publishing Company, 1934.

SOCIAL ATTRACTION PATTERNS BETWEEN ELEMENTARY-SCHOOL CHILDREN AND STUDENT-TEACHERS: SOCIOMETRIC ANALYSIS[1]

WALTER D. SMITH

Department of Psychology
Florida State University

This study is an attempt to define and interpret the developing patterns of social attraction existing between 184 elementary school pupils (nursery through grade six) and fifty-two student-teachers. These students were fulfilling certain requirements for the teacher's certificate by practicing one college semester under the supervision of a critic teacher. By using sociometric techniques similar in design to those introduced by J. L. Moreno,[2] patterns of attraction and repulsion are revealed in their initial stages of development. An appraisal of the critic teachers' awareness of the social acceptability of their student-teachers by the pupils provides data which have many implications for teacher evaluation and teacher training.

More specifically the purposes of the study are stated in the form of questions concerning each problem or relationship to be investigated:

1) Do children in the nursery, kindergarten and elementary school groups evidence consistent patterns of choice and rejection for student-teachers who are acting in a supervisory capacity with the children? Do children show strong and consistent differentiation in choosing among the student-teachers? This may demonstrate that not 'just any' adult is to be given training as an elementary school teacher, that children do distinguish

---

[1] This report is a portion of the writer's doctoral dissertation, Social Attraction between Elementary School Children and Student Teachers. *Microfilm Abstracts,* Vol. 10, No. 4, Publication No. 2008, Ann Arbor: University Microfilms, 1950. Acknowledgment is made to Dr. Willard C. Olson, University of Michigan, who directed the original research and to the Graduate Research Council, Florida State University, whose grant has made continuance of the study possible.

[2] J. L. Moreno, *Who Shall Survive?* Washington, D. C.: Nervous and Mental Disease Publishing Company, 1943.

consistently between those whose presence and association is pleasant and desirable, even in nursery school groups.

2) Do children tend to show greater differentiation between criteria of the sociometric test as they become better acquainted with student-teachers and learn the special skills and abilities of each, and do they tend to choose each individual for the situation in which he performs best? Do inter-criterial[3] correlation coefficients between 'play' and 'work' choices decrease with each additional month in the scores received on both choice and rejection?

3) Is a student-teacher's adjustment in his group of co-workers related to his adjustment in the children's groups? Is acceptance or rejection by one group indicative of acceptance or rejection by the other?

4) Due to different concepts of values and good human relations between adults and children, are critic teachers able to perceive accurately the true nature of the student-teacher-pupil relations? Will their predictions or estimates of children's choices and rejections for student-teachers be similar to the actual situation? Will their awareness or perception improve with additional time of acquaintance?

Experimental situation.—The subjects of the study were the pupils, student-teachers, and critic teachers (or classroom teachers) of the University Elementary School of the University of Michigan. Within this school various research projects were carried out from time to time by the research staff in child development.

The school was comprised of nine grades or groups: Nursery I, Nursery II, Kindergarten, and Grades I through VI, each under the direction of a regular teacher who also served as critic teacher in the student-teacher training program. These groups ranged in numbers from thirteen to twenty-four and were generally equally divided according to sex. It will be noted that the school served two main purposes—that of a research unit and a teacher training unit.

The child population of the study appears to have been unusual with higher socio-economic background than found in most public schools and considerable superiority to the general population in intelligence (Stanford-Binet IQs ranged from 79 to 196 with few falling below 110). Similarly the student-teacher population (fifty-one females and one male) appeared to have been a select one in terms of socio-economic background, although background data could not be accurately evaluated because of their sometimes ambiguous nature. Thirty-seven of this group were twenty-one years of age; the age range was nineteen to twenty-eight years. In no case had any student-teacher had previous teaching experience.

---

[3] The term, *inter-criterial,* is used here to represent the relationship between the two criteria, *play* and *work.*

Student-teaching assignment.—Room and grade assignments of student-teachers were generally based on preference expressed prior to the beginning of the school term. All assignments were made for an entire semester. With his critic teacher each student-teacher planned his work schedule in such a way that he would spend twelve and one-half hours in his assigned room each week. The number assigned to the several grades varied between four and seven. A considerable amount of overlapping of schedules and regular weekly meetings with the critic teacher served to acquaint the student-teachers with each other in a few short weeks. Generally critic teachers attempted to rotate each student-teacher's duties in order that each would be able to teach or work in every area of the school program.

Classroom and critic teachers.—The classroom teachers, or critic teachers, were teachers with several years of teaching experience. Five of this group had the master's degree and others were doing work towards that degree. The ages of all critic teachers were under thirty-five. The sixth-grade teacher was the only male.

Materials and instruments used.—Using criteria possessing equal meaning for children at all grade levels studied several types of sociometric tests were derived. These tests, similar in design to the usual device as described by Northway,[4] Jennings,[5] and Bronfenbrenner,[6] proposed to explore and measure lines or pathways of attraction and repulsion in two channels: child-to-student-teacher and student-teacher-to-student-teacher.

---

[4] Mary L. Northway, "A method for depicting social relationships by sociometric testing," *Sociometry*, III, 2(April, 1940), 144–50.

[5] Helen H. Jennings, *Leadership and Isolation.* New York: Longmans, Green and Company, 1943.

[6] Urie Bronfenbrenner, *The Measurement of Sociometric Status, Structure, and Development.* Sociometry Monographs, No. 6. New York: Beacon House, 1945.

(1) Form A, administered to all children:

"Tell me which of the student-teachers you would like *best* to have play with you any of the games that we play," and "Tell me which of the student-teachers you would like *best* to show you how to do things and to help you when you are working, or making something, or studying—just any of the work that we do."

The negative sociometric questions were similar to the positive ones above except the word 'least' was substituted in each case for 'best.'

Some variations in instructions occurred in the upper elementary groups where each child was asked to write his own response. Two spaces for names were left after each question.

(2) Form B, administered to all student-teachers:

To investigate student-teachers' choices for each other they were given these instructions: "Please write the name of the student teacher in your room whom you would like *best* as a companion in your leisure-time activities. Perhaps there are more than one. Name those you really prefer," and "Write the name of the student-teacher in your room whom you would like *best* as a co-worker in presenting any unit or work activity to the class." Spaces for two names were left under each question. In order to make the test as meaningful as it had been for the children, some changes from the children's tests were made in wording.

(3) Form C, administered to all critic teachers:

Critic teachers were asked to rank student-teachers, on the basis of their knowledge or awareness of pupil-student-teacher relationships, as they thought the children's choices would rank them. Rankings on both criteria, 'play' and 'work,' were made.

Test administration.—The larger task of testing was delegated to the regular classroom teachers. All tests administered to children from the nursery groups through grade three required individual testing; all other tests were administered on a group basis. Student-teachers completed their sociometric tests at their weekly meeting following most closely the date of administration of the children's tests.

Time interval between tests.—Because a major purpose of the study was to catch the picture of developing relations within each of the nine groups and to indicate the presence of any characteristic trends, the necessity of administering sociometric tests at brief intervals becomes apparent. The logic of this procedure and the appropriate time intervals have been discussed by Moreno[7] and Jennings.[8]

The time interval decided upon was four weeks. The period under study was limited to the three months beginning with the opening of school in September, 1949, and extending to December 12, 1949. Short intervals appear satisfactory when the group members are not well acquainted at the initial test, making it imperative that each individual make some new adjustment to accomodate each of his new associates (Moreno[9]).

The first test occurred during the fourth week of acquaintance. By this arrangement children and student-teachers had a short period to become acquainted and draw first impressions; the final tests show the durability of these early impressions. The tests are then labelled according to the date of administration as October 17, November 14, and December 12.

Use of identifying pictures.—To assist the children in making choices a photograph of each student-teacher was made and the pictures (3 × 4 inches) were attached in a circular arrangement to a plain cardboard (18 × 18 inches). Each child could then identify and choose the preferred person's picture.

II. UTILIZATION OF SOCIOMETRIC DATA

Tabulation of data.—All sociometric data from each group on each test were tabulated according to the matrix design explained by Bronfenbrenner,[10] Northway and Potashin,[11] and Jennings.[12]

---

[7] Moreno, *op. cit.*

[8] Helen H. Jennings, "Sociometry in action," *Survey Midmonthly*, LXXXIV, 2(February, 1948), 41–44.

[9] Moreno, *op. cit.*

[10] Bronfenbrenner, *op. cit.*

[11] Mary L. Northway and Reva Potashin, "Instructions for using the sociometric test," *Personality and Sociometric Status.* Sociometry Monographs, No. 11, New York: Beacon House, 1947, 67–71.

[12] Helen H. Jennings, *Sociometry in Group Relations.* Washington, D. C.: American Council on Education, 1948.

Combining scores from work and play choices.—Due to the unwieldy nature of scores representing several segments of the classroom social climate and to the need for single measures to be used in comparing social status with various aspects of personality and performance, the combining of scores on different criteria with a single score obtaining which would represent a composite picture of social status seemed warranted. This procedure has been followed by Bonney[13] and Northway and Potashin.[14]

At the same time the correlations between scores on the various criteria are usually high and statistically significant, thus rendering the use of composite scores a statistically sound procedure (Horst[15]).

Weighting of scores.—Bronfenbrenner[16] allowed three choices on each criterion and dispensed entirely with any weighting system giving all choices equal weight as have others (Newstetter, Feldstein, and Newcomb[17]). Other investigators have quite arbitrarily weighted scores in terms of priority of preference (Bonney,[18] Northway[19]). The evidence would indicate that a weighting technique provides no advantage over unweighted scores and that the latter technique would lend itself admirably to the present study. The raw score for each individual was the number of choices received.

III. STATISTICAL ANALYSIS OF SOCIOMETRIC DATA

Constancy of choices for student-teachers.—Will certain adults quickly attract children and maintain this attraction (or choice status) while others are consistent in being 'neglected' or 'rejected' by the children in the choosing situation? Considerable variability between groups and within groups was found with Pearson r's ranging from −.801 (Nursery I, October–December choice status) to +.857 (Grade V, November–December choice status). However, by combining r's from all groups by the r to z transformation recommended by Fisher[20] for small samples, it is found that the estimated total r's thus obtained do attain statistical significance (Table 1). Levels of significance were read directly from Fisher's Table V-A.[21]

---

[13] Merl E. Bonney, "The constancy of sociometric scores and their relationship to teacher judgments of social success, and to personality self-ratings," *Sociometry*, VI, 4(November, 1943), 410.

[14] Northway and Potashin, *op. cit.*, p. 70.

[15] Paul Horst, *The Prediction of Personal Adjustment*. New York: Social Science Research Council, Bulletin 48, 1941.

[16] Bronfenbrenner, *op. cit.*

[17] W. I. Newstetter, N. J. Feldstein, and T. M. Newcomb, *Group Adjustment: A Study in Experimental Sociology*. Cleveland: School of Applied Social Sciences, Western Reserve University, 1938.

[18] Bonney, *op. cit.*

[19] Mary L. Northway, "Outsiders, a study of the personality patterns of children least acceptable to their age mates," *Sociometry*, VII, 1(February, 1944), 11.

TABLE 1.—ESTIMATED TOTAL CORRELATION COEFFICIENTS BETWEEN STUDENT-TEACHERS' TOTAL CHOICE SCORES (PLAY AND WORK COMBINED) FROM CHILDREN ON THREE TESTS

$n = 52$		Oct.–Nov.	Nov.–Dec.	Oct.–Dec.
Estimated "r"	=	.509	.594	.460
df	=	34	34	34
p	=	$<.01$	$<.01$	$<.01$

Assuming that the several samples comprising each of these estimated r's are drawn from equally correlated populations, one may conclude that a marked relationship exists in the status of student-teachers in children's groups from time to time; the student-teacher who is accepted at one time would tend to be accepted at another time.

Constancy of rejection for student-teachers.—"Do student-teachers who are once rejected by children remain rejected?" Estimated r's (Table 2) were again calculated. Each r was found to differ significantly from zero at the .01 level indicating a considerable constancy in the children's rejection of student-teachers, an even greater constancy than was demonstrated in student-teachers' choice status. Rejection by children on one occasion is highly indicative of rejection at another point in time.

Differentiation of play and work choices.—"Is there a significant relationship between 'play' and 'work' status?" Choice scores

---

[20] Quinn McNemar, *Psychological Statistics*, pp. 348–49. New York: John Wiley and Sons, Inc., 1949.

[21] Ronald H. Fisher, *Statistical Methods for Research Workers*, p. 202, 8th ed. Edinburgh: Oliver and Boyd, 1941.

TABLE 2.—ESTIMATED TOTAL CORRELATION COEFFICIENTS BETWEEN STUDENT-TEACHERS' TOTAL REJECTION SCORES (FROM CHILDREN) ON THE THREE TESTS

n = 52		Oct.–Nov.	Nov.–Dec.	Oct.–Dec.
Estimated "r"	=	.521	.716	.653
df	=	34	34	34
p	=	<.01	<.01	<.01

received from children on the two criteria were correlated yielding the data of Table 3 which represent the combined r's from the nine groups (combined by the r to z transformation).

TABLE 3.—ESTIMATED TOTAL CORRELATION COEFFICIENTS BETWEEN STUDENT-TEACHERS' 'PLAY' AND 'WORK' CHOICES (FROM CHILDREN) ON THE THREE TESTS

n = 52		October	November	December
Estimated "r"	=	.211	.321	.435
df	=	34	34	34
p	=	>.10	>.05	<.01

The increasing values of r provide evidence that some influence similar to halo effect became more and more operative with each additional month of acquaintance. An r-value in December of .435, significant at the .01 level, indicates that children's choice or acceptance of student-teachers on one criterion becomes indicative of choice on the other criterion when adequate time for crystallization of preferences is allowed.

Differentiation of play and work rejection.—Values of estimated r were obtained between play and work rejection on each of the three tests (children's rejections for student-teachers.) Table 4

TABLE 4.—ESTIMATED TOTAL CORRELATION COEFFICIENTS BETWEEN STUDENT TEACHERS' 'PLAY' AND 'WORK' REJECTION SCORES (FROM CHILDREN) ON THE THREE TESTS

n = 52		October	November	December
Estimated "r"	=	.304	.655	.741
df	=	34	34	34
p	=	>.05	<.01	<.01

shows that the relationship between the criteria of rejection is even greater than that between the criteria of choice. The values of r rapidly gain statistical significance and an r of .741 in December, significant at the .01 level, provides rather conclusive evidence of a general factor running throughout the test. The pessimistic conclusion suggested is that if the student-teacher is rejected by children in one situation, he is likely to be rejected in other situations. This tendency becomes more marked with each additional month as group structures become more set and greater unanimity of opinion is reached on the matter of the most rejected individuals.

Within a few brief weeks children appear to decide which student-teachers are least preferred and tend to consistently name them on all criteria. This tendency is even more pronounced than is the tendency for each student-teacher to maintain similar choice status on the criteria employed over the three months' interval. Factors which contribute to rejection may be more generalized in their effect on student-teacher-pupil relations and more stable and enduring than those factors contributing to choice expressions.

Relationship between student-teachers' choices received from student-teachers and from children.—Whether student-teachers who are well-adjusted socially in their own groups, adjustment being defined by high social status scores, are also well-adjusted in the children's groups may be answered by determining the nature of the relationship existing between the choice and rejection scores garnered in the two situations. While a more comprehensive analysis of each individual's sphere of social contacts would perhaps lead to greater insight into this relationship, a statistical analysis may serve to uncover certain group trends.

Correlation coefficients between these two social status measures have been obtained for each of the nine groups of student-teachers; again r's are combined and estimated total r's obtained for each month. An inspection of Table 5 leads to the conclusion that social status of the student-teachers in their own groups is not related in any easily generalized manner to their status in the children's groups. Whether these findings are a function of the small samples studied or whether the trend would continue when larger samples become the center of investigation leaves room for conjecture.

TABLE 5.—ESTIMATED TOTAL CORRELATION COEFFICIENTS BETWEEN STUDENT-TEACHERS' CHOICE SCORES RECEIVED FROM CHILDREN AND THOSE RECEIVED FROM THE OTHER STUDENT-TEACHERS IN THE GROUP

n = 52		October	November	December
Estimated "r"	=	.279	.237	−.316
df	=	34	34	34
p	=	>.05	>.10	>.05

Relationship between student-teachers' rejections received from student-teachers and from children.—A comparison of negative sociometric data (rejections) in a similar fashion to the preceeding choice comparison yields the data of Table 6; positive values of estimated total r are evidenced each month. Only a moderate relationship between the extent of rejection in children's groups and rejection in student-teachers' groups is revealed. However, in these low positive values a trend is suggested toward a greater relationship with each additional month; by December the r-value of .357 has reached the .05 level of significance.

TABLE 6.—ESTIMATED TOTAL CORRELATION COEFFICIENTS BETWEEN STUDENT-TEACHERS' REJECTION SCORES RECEIVED FROM CHILDREN AND FROM STUDENT-TEACHERS

n = 52		October	November	December
Estimated "r"	=	.211	.302	.357
df	=	34	34	34
p	=	>.10	>.05	<.05

It may be concluded that within the limits of the present study little uniform relationship exists between the way a student-teacher is accepted by his adult associates and the way children accept him. However, it would appear that a more direct and positive relationship develops between these two measures over a period of time; the data suggest that a greater time limit than three months is required to test this hypothesis. Further, it might be surmised that the children's system of evaluation of adult associates is vastly different from the adult's system of evaluation of adults. Social acceptance by adult associates offers no assurance that children will entertain similar feelings. These findings would prompt a constant alertness on the part of the person in a supervisory and administrative capacity in order to see teachers as children view them.

Critic teachers' awareness of 'most chosen' and 'least chosen' student-teachers (chosen by children).—For the forward-looking critic teacher who would improve the interpersonal relations between student-teachers and pupils, a knowledge of those relationships becomes imperative. As one aspect of the analysis of pupil-student-teacher relations in the classroom, critic teachers were asked in advance of each sociometric testing period to rank student-teachers on the two criteria ('play' and 'work') as they thought the children's choices would rank them.

While a knowledge of social status of group members might be of great benefit to one attempting to improve human relations in the classroom, some evidence exists which would indicate that a very accurate awareness by the individual member is not to be expected. It has been pointed out by Moreno[22] that the individual member is only partially aware of his position. Newstetter, Feldstein, and Newcomb[23] have reported a mean correlation coefficient (r) of .756 ≠ .20 between camp counselor estimates of a child's acceptance and the obtained index of group status.

Fauquier and Gilchrist[24] have suggested that some of the qualities of leadership in children's groups—dominance, aggression, boldness, impulsiveness, excitability, and alertness—frequently serve to hinder the adult's recognition of the group control resting in these individuals. Because the basis for judgment of choice is different for children and adults, critic teachers might experience some difficulty in making accurate estimates of any student-teacher's acceptance by children.

While it is unlikely that critic teachers can estimate with much accuracy children's choices for the 'average group' in acceptance, an awareness of extreme cases (those who are 'most chosen' and those who are 'least chosen' by the children) does lie within the realm of reasonable expectation. Detection of the extreme cases on the lower end of the distribution of children's choices is particularly important.

---

[22] Moreno, *op. cit.*, p. 339.

[23] Newstetter, Feldstein, and Newcomb, *op. cit.*

[24] William Fauquier and John Gilchrist, "Some aspects of leadership in an institution," *Child Development*, XIII, 1 (March, 1942), 55–64.

In appraising teachers' judgments, it was required that each judgment meet with an arbitrarily designed criterion of correctness. The criterion of correctness was that the student-teacher who actually obtained the most choices on each criterion ('play' and 'work') would be ranked in either first or second place in the group while the student-teacher who received the least number of choices would be ranked either lowest or in next to the lowest position in the group.

The values in Table 7 show a picture of decreasing awareness; the high values of sixty-seven per cent in October for both extreme groups indicate that the critic teachers were rather quick to observe and understand the children's reactions to their several student-teachers. The steadily decreasing values from month to month indicate that the complexity of the pupil-student-teacher ties defied comprehension by the critic teachers. Moreno explains that "the intricacies of the children's own associations prevent the teacher from having a true insight," and adds that "this appears as one of the great handicaps in the development of teacher-child relationships."[25]

TABLE 7.—PER CENT OF CORRECT ESTIMATES BY CRITIC TEACHERS OF 'LEAST CHOSEN' AND 'MOST CHOSEN' STUDENT-TEACHERS

October		November		December	
Highest	Lowest	Highest	Lowest	Highest	Lowest
67	67	61	61	50	33

It may be concluded that the critic teachers were initially able to pick-out with considerable accuracy the 'most chosen' and the 'least chosen' student-teachers in their groups and that the accuracy with which this process was carried out decreases with additional increments of time.

---

[25] Moreno, *op. cit.*, p. 54.

V. SUMMARY AND CONCLUSIONS

An investigation is reported of the nature of social attraction and repulsion directed from 184 elementary school children to fifty-two student-teachers and of patterns of social attraction and repulsion between the student-teachers by sociometric tests administered to the children and the student-teachers at monthly intervals, October to December, 1949. The children and students recorded their choices and rejections for associates for 'play' and 'work.' Relationships between aspects of social status were determined; critic teachers' awareness of student-teachers' acceptance by children was revealed. The findings were:

1) Over a period of three months student-teachers' acceptance by children remained remarkably constant. Acceptance at one time is indicative of later acceptance; rejection on one occasion tends to be even more indicative of rejection on later occasions. However, shifts in status occur in sufficient frequency and amount in all grades to warrant efforts for improvement.

2) Increasingly high relationships in student-teachers' choice and rejection scores for play and work (inter-criterial) from children in successive months suggest either a greater breadth of acquaintance or an influence in the choice process similar to halo effect.

3) The acceptance of student-teachers in their own groups (of student-teachers) showed no significant relationship to their acceptance in the children's groups. However, the data suggest an increasing positive relationship from month to month.

4) Critic teachers initially displayed considerable awareness concerning the status of the 'highest chosen' and the 'least chosen' student-teachers. Judgments appeared to grow less accurate with each passing month.

The procedures employed are recommended as diagnostic techniques for use in exploring student-teachers' relationships with children. The notable inadequacies of rating scales and administrative opinion as to the success of teachers in their relationships with children prompt the investigator to turn to the children for this answer.

BOOK REVIEWS

EDWARD A. STRECKER, FRANKLIN C. EBAUGH, JACK R. EWALT AND LEO KANNER. *Practical Clinical Psychiatry.* New York: The Blakiston Company, 1950, pp. 506.

For some time to come it is the general medical practitioner who will treat the vast majority of psychiatric patients, particularly those who are suffering from psychoneurotic and psychosomatic disorders. With this fact in mind, the present or seventh edition of a popular textbook in psychiatry since 1925 is written to satisfy the needs of medical students and doctors in general practice. The content is presented in fifteen chapters in the usual textbook fashion. Most of the chapters are devoted to the consideration of special psychoses or neuroses. These include "Organic Reaction Types," "Toxic Psychosis," "Affective Reaction Types," "Schizophrenic Reaction Types," "Constitutional Psychopathic Inferior," "Reactions of Developmental and Constitutional Defects," "Paranoid Reaction Types and Paranoia," "Traumatic Reactions," and "Psychoneuroses."

The first chapter is on "Personality Development and Function" and here we get an early expressed appreciation of Adolf Meyer's orientation. To Dr. Meyer this volume is dedicated; however, the chapter does contain an expressed appreciation of the fact that a large segment of psychiatric thinking of today is derived from psychoanalysis. In this chapter particularly, and here and there in other chapters dealing with matters other than interpretive material, there is a tendency to use short quotations which seem, at times, to be in contrast with more rigorous thinking. Hence, we find a quotation in this chapter, for example, from Lin Yutang which reads as follows: "Our mind was originally an organ for sensing danger and preserving life. That this mind eventually came to appreciate logic and correct mathematical equation I consider a mere accident. Certainly it was not created for that purpose. It was created for sniffing food, and if after sniffing food it can also sniff an abstract mathematical formula, that's all to the good." The section on psychopathology is nearly entirely limited to a brief description of Freudian mental mechanisms. The treatment of "Methods of Psychiatric Examination" is quite detailed but without enough explanation and illustration of use of the detail. The chapter on "Support

Psychotherapy," which particularly features four 'helps,' does contain much that can be useful. "Support Psychotherapy" is distinguished from psychoanalysis by the authors and is referred to as 'target therapy' since its main purpose is to unearth as completely as possible the hidden psychopathology. The implication is that 'support psychotherapy' can be relevant, useful, quite thorough at times. Skillfully used, they say, it almost always traces the main outlines of underlying mental conflict and sometimes it penetrates quite deeply. One piece of advice in this connection is relevant; namely, that the psychotherapists' maturity should incline him to freely utilize even measures which he somewhat dislikes if they are needed to help his patients. Non-directive counselling is described as too rigid for everyday use at times. 'I' of one author appears at times to get mixed with the 'we' of all collaborators. For example, "I have come to believe that the development of what might be called a normal amount of anxiety is a basic need of childhood"; also, "Finally, I do believe that something that might be called spiritual outlook, perhaps with some religious values and convictions, is an important constituent basic need." Which 'I' of the authors this is, it doesn't say; the chapter as a whole is written in 'we' terms. Group therapy is given lip service to or is praised although only a short paragraph is devoted to this form of therapy in "Support Psychotherapy."

The last chapter on the psychopathology of childhood is written by Dr. Leo Kanner and is a compact and good chapter for the purpose. All in all, this seventh edition of a popular textbook is likely to be as well received as have been its predecessors and deservedly so in spite of the fact that one could criticize it for not containing more psychodynamic content or emphasis. Even though the book is written with the medical practitioners in mind, this is intended as a first textbook and should be evaluated as such. As such it is, as its preceding editions, a good, substantial text.

H. Meltzer

Psychological Service Center
St. Louis, Missouri

Horace B. English. *Child Psychology*, New York, Henry Holt and Company, 1951, pp. 561.

Many psychological text books are little more than summaries of one experiment after another. These may be pieced together, more or less successfully, by some sort of logical framework, but the reader comes away with the feeling of having spent his time poring over psychological abstracts.

English's *Child Psychology* belongs in a different category. He states frankly that he has a deep concern for children, and believes that we are daily discovering through science ways of promoting their welfare. His book was admittedly written for the purpose of getting those ways adopted as widely as possible. In accordance with this aim, English has taken the pains to develop an easy informal style, and has produced a book which is preëminently readable.

The book is addressed, not to the research worker, but to the young prospective teacher, who may possess only a limited background in psychology. It should prove usable as a text in elementary child psychology.

The entire book is centered around the project of making a case study of a child. Not content with merely furnishing the outline for such a project, English gives detailed instructions about how to make contacts, what to say to parents, how to interview the child, how to observe him in school, how to arrange to take him to a movie, how to make records (carefully distinguishing fact from interpretation), and how to prepare the final report. A very complete case study is included in the book to serve as a sample.

Although the beginning student may see only the practical side of the book, anyone familiar with research in the field will become progressively aware that the various opinions are consistent with experimental conclusions. In other words, English has given his readers the results of a lifetime of study without inflicting upon beginning students a tedious and (for them) unnecessary mass of detail.

Melvin G. Rigg

New Mexico Highlands University

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 44 March, 1953 Number 3

EFFECTS OF DIRECTIONS REGARDING GUESSING ON ITEM STATISTICS OF A MULTIPLE-CHOICE VOCABULARY TEST

FRANCES SWINEFORD and PETER M. MILLER

Educational Testing Service
Princeton, N. J.

When an examinee meets a test item which is totally unfamiliar to him, he may omit the item or he may make a wild guess. There is no such thing for him as a 'shrewd guess' when the item is completely divorced from his previous experience. If only an occasional examinee finds an occasional item of this nature, nothing he does can seriously affect either his score or the descriptive statistics of the test or items. If, on the other hand, a test were to contain a substantial number of items totally unfamiliar to a large proportion of the examinees, there might be a noticeable effect on test scores, test statistics, and item statistics. One of the aims of this study is to investigate such a situation.

It is a further purpose of this study to investigate the amount of guessing that is likely to occur under different instructions to the examinee, to find what relationship may exist between amount of guessing and performance in the area covered by the test, and to determine the effects of guessing on various statistics.

In order to accomplish these purposes, a 100-item vocabulary test was constructed, containing eighty regular items of appropriate difficulty for the group to be tested, ten extremely difficult items, and ten nonsense items. The extremely difficult items contained stem words which appear in *Webster's New Collegiate Dictionary* but which were unfamiliar to the writers and to their colleagues to whom the list of words was submitted; it is unlikely that any of these words would be familiar to the average college graduate. The nonsense items, which were 'scored' by an arbitrary, randomly devised key, contained stem 'words' which are not words at all. The items were not arranged in order of difficulty. The twenty special items were distributed at random throughout the test. They are listed below:

Special Items

Difficult items:

11. MURICATE: 1-prickly 2-closed 3-purple 4-emblazoned 5-enfeebled
20. NUBILE: 1-marriageable 2-black 3-oily 4-predatory 5-removable
30. CONDIGN: 1-elevated 2-deserved 3-taciturn 4-fabricate 5-supplicate
34. HIRCINE: 1-epicene 2-placid 3-multifarious 4-goatlike 5-enslaved
53. LANNER: 1-rope 2-specialty 3-wrench 4-falcon 5-schooner
57. DECORTICATE: 1-husk 2-testify 3-swear 4-cleanse 5-affirm
59. GESTION: 1-animadversion 2-hint 3-altercation 4-humor 5-conduct
76. SAPID: 1-ailing 2-innocuous 3-palatable 4-uninformed 5-aged
78. JINK: 1-coin 2-dodge 3-crack 4-dance 5-jot
84. FAINAIGUE: 1-rumpus 2-mausoleum 3-cheat 4-invalidate 5-conspiracy

Nonsense items:

17. QUINTULENT: *1*-feverish 2-faltering 3-acrid 4-prurient 5-decaying
27. TAMORIN: 1-shrew 2-drum 3-woodpecker *4*-urn 5-associate
38. ARDICIAN: 1-handyman *2*-suitor 3-barber 4-geologist 5-undertaker
44. PALIENT: *1*-dim 2-important 3-sharp 4-twinkling 5-friendly
50. VENTRESCULATION: 1-wound 2-eloquence *3*-latticework 4-window 5-profanity
60. HILN: 1-cupboard 2-handle *3*-meadow 4-outhouse 5-relation
65. RHUSTATE: 1-inflamed 2-stopped up 3-frustrated *4*-reddish 5-ill-bred
69. BRUNNAGE: 1-fog 2-anger 3-rigging 4-darkness *5*-effrontery
72. SUSCERN: 1-see 2-dissociate *3*-suspect 4-worry 5-repudiate
88. WALDER: 1-meander 2-lancer 3-renege 4-mason *5*-mend

Note: The most popular responses to the nonsense items are as follows: 17(2); 27(4); 38(1); 44(1); 50(2); 60(2); 65(4); 69(5); 72(3); 88(1).

The test was administered in three forms, which differed only in the cover-page instructions. Following detailed directions for proper use of the separate answer sheet, directions which were the same on all forms, the remaining directions were as follows:

Form 1

> It is not expected that everyone will finish in the time allowed. Do not hurry, but work steadily and as quickly as you can without sacrificing accuracy.
>
> You will be given 30 minutes to work on this test.

Form 2

> It is not expected that everyone will finish in the time allowed. Do not hurry, but work steadily and as quickly as you can without sacrificing accuracy.
>
> You are advised to answer a question only if you are sure of the answer. You should *not* guess since wrong answers will result in a subtraction from the number of your correct answers.
>
> You will be given 30 minutes to work on this test.

Form 3

> It is not expected that everyone will finish in the time allowed. Do not hurry, but work steadily and as quickly as you can without sacrificing accuracy.
>
> Answer all questions about which you have any knowledge. You are advised, further, to make a guess on unfamiliar words: a shrewd guess is more often right than wrong. Your score on this test will be based on the number of your correct answers; no deduction will be made for wrong guesses.
>
> You will be given 30 minutes to work on this test.

The three forms were distributed in such a way that every third examinee received the same form. No mention was made of the fact that there were different forms, and it is unlikely that the examinees were aware of that fact before they had an opportunity, later, for discussions with one another. This method of distributing the forms among as many as eight hundred examinees virtually assures samples that are equivalent for all practical purposes. We shall be interested only in experimental differences which are significantly greater than the sampling differences that are likely to arise.

When this study was proposed, three hypotheses were set down; namely,

1) There will occur some guessing on the part of all three groups.

2) The amount of guessing will vary only slightly with differences in instructions.

3) There is little or no relationship, either positive or negative, between ability and the tendency to guess.

No hypotheses were offered concerning the test statistics. These will be examined, however.

Seven scores were obtained on each paper. They are:

A. Score (R) on eighty regular items.
B. Score (R) on ten difficult items.
C. 'Score' (R) on ten nonsense items.
D. Total score (A + B + C).
E. Number of responses to ten difficult items.
F. Number of responses to ten nonsense items.
G. Total number of responses to special items (E + F).

There were 267 cases who took each form, 801 cases in all. In Table 1 are listed the means and standard deviations of the seven scores.[1]

The means for Score A (eighty regular items) are remarkably similar. Although the mean for Form 2 (do not guess) is the smallest and that for Form 3 (try all items) is the largest, as would be expected from the nature of the directions, their difference of less than two score points is both unimportant practically and no greater than one which might reasonably occur by chance.

Taking into consideration the number of difficult items attempted (Score E), we find the means for Score B (ten difficult items) to be extremely close to the expected chance means of 1.84, 1.10, and 1.91. The significant differences for Score B between Form 2 and the other two forms are, therefore, related to the number of items marked rather than to knowledge of the words themselves.

A curious result occurs in connection with Score C (ten nonsense items). Each of the obtained means, 2.48, 1.63, and 2.69, is significantly greater than the corresponding expected chance figure: 1.84, 1.10, and 1.90. Thus the examinees tend to agree with each other and with the arbitrary key to an extent which cannot be explained by chance alone. To test the possibility that the selection of particular responses by the examinees might

---

[1] Should the reader be interested in the score distributions, they may be obtained from the authors at Educational Testing Service, Princeton, New Jersey.

TABLE 1.—MEANS AND STANDARD DEVIATIONS OF SEVEN SCORES ON EACH FORM

Score	Form 1		Form 2		Form 3	
	Mean	SD	Mean	SD	Mean	SD
Number right:						
A (80 regular items).....	42.87	11.33	41.25	11.34	42.91	10.70
B (10 difficult items)	1.83	1.31	1.03	1.11	1.96	1.34
C (10 nonsense items)....	2.48	1.36	1.63	1.56	2.69	1.44
D (A + B + C)........	47.18	11.94	43.92	12.06	47.56	11.37
Number of responses:						
E (10 difficult items)....	9.18	2.16	5.52	4.10	9.54	1.59
F (10 nonsense items)...	9.16	2.19	5.48	4.13	9.50	1.67
G (E + F).	18.34	4.28	11.00	8.14	19.04	3.19

be associated with the ability measured by the eighty regular items (Score A), the correlation between Scores A and C was computed for the Form 3 group. The correlation, .105, does not differ significantly from zero. (Item biserial correlations, to be presented later, constitute another expression of the same finding.) We may conclude, therefore, that the arbitrary key happened to include some unusually attractive keyed answers which were almost equally attractive to examinees of all levels of ability as measured by Score A. This being so, another arbitrary key might equally well have produced mean scores that are substantially under 'chance' values if it did not include any of the popular choices. (The most popular alternatives have been indicated below the set of items given earlier in this report.) Apart from the unexpectedly high values, the Score C means bear the same relationships to one another as do the Score B means. In each instance the mean for Form 2 is lower than the other two means by statistically significant amounts.[2]

---

[2] Group comparisons for Scores A, B, and C have been made by the method of analysis of variance. The F ratios are 1.91, 42.2, and 39.3, respectively. With 2 and 798 degrees of freedom in each instance it is clear that the between groups variance is not significant for Score A but is significant beyond any doubt for Scores B and C.

On Score D (score on all one hundred items) the means for the Form 1 and the Form 3 groups are very similar. Their difference of 0.38 divided by its standard error, 1.01, is only 0.38. The mean for the Form 2 group, however, is less than the other two means by amounts which are in each case more than three times the standard error of the difference and, hence, which are statistically significant beyond the one per cent level of confidence.

Scores E (number of responses to ten difficult items) and F (number of responses to ten nonsense items) are so similar that they can be considered together. For both Form 1 and Form 3 the mean number of items tried out of each set of ten items is more than 9.00. Nevertheless, the small difference of less than one-half of an item between these groups is statistically significant at the five per cent level of confidence. The Form 2 group, on the other hand, followed the instructions not to guess to the extent that the mean number of items answered is about 5.5 for each set of items. This figure, however, by no means represents the typical Form 2 examinee. Despite the instructions, nearly forty per cent of the students tried all the special items. An equal number tried no more than twenty-five per cent of the items, whereas the rest are scattered through the remaining score range. In contrast, about eighty per cent of the Form 1 and Form 3 groups tried all the special items, and no secondary mode appears at the low end of the distributions.

The means for Score G (E + F), of course, add no new information, but from the standard deviations of E, F, and G there can be calculated the correlations, r_{EF}, which are .932 for Form 1, .954 for Form 2, and .924 for Form 3. Correlations of this magnitude, based on so small a number of items, indicate an unusually high degree of consistency of response. Reaction to the difficult items was essentially the same as reaction to the nonsense items. It should be noted that the correlations are somewhat inflated by the large numbers of examinees who tried all the special items. If all such cases are omitted from the calculations, however, values of .87, .83, and .85 are obtained, which may properly be regarded as lower bounds of the relationship that actually exists. This extremely consistent behavior appears from the present data to be in the nature of a 'response set,' determined in part by the test directions and in part by the personality or previously established habits of the examinee.

The relationship between ability in the area of the test and this tendency to respond to the special items can be measured by the correlations between Score A and Score G. These are .039 for Form 1; .160 for Form 2; and .007 for Form 3. The Form 2 correlation, although too low to be of practical importance, is statistically significant at the one per cent level of confidence. In other words, when instructions were given not to guess, there was a slight but real tendency for the more able students to disregard the instructions and to attempt the special items. The Form 1 and Form 3 correlations can be regarded as chance deviations from zero.

TABLE 2.—FREQUENCY DISTRIBUTIONS OF NUMBER OF OMISSIONS FOR EACH ITEM

Number of Omissions	Form 1		Form 2		Form 3	
	Regular Items	Special Items	Regular Items	Special Items	Regular Items	Special Items
140–149..				1		
130–139..				4		
120–129..			1	8		
110–119..			1	1		
100–109..			3	5		
90– 99..			—	1		
80– 89..			1			
70– 79..			5			
60– 69..			4			
50– 59..			7			
40– 49..			5			
30– 39..			14			
20– 29..	3	13	8		1	
10– 19..	12	7	9		9	18
0– 9..	65		22		70	2
Total..	80	20	80	20	80	20
Mean....	6.8	21.0	34.8	120.5	5.9	13.5

The effect of the different instructions on examinee behavior is strikingly portrayed in another fashion in Table 2, which gives frequency distributions of the number of omissions per item. (An unanswered item is considered an omission only if a later item has been answered. Otherwise it is classified as 'not reached.') Similarities and differences are so clear-cut that tests of statistical significance are not necessary. Under all the conditions the special items were omitted more frequently than the regular, easier items, but the difference between special and regular items is particularly great for Form 2 (instructions not to guess). Only a few more omissions occur for Form 1 (no instructions on guessing) than for Form 3 (instructions to answer every item).

If the different instructions affect the typical examinee's rate of work, proper timing under various conditions becomes an important problem. The use of a speeded test would be necessary to establish typical rates of work. In the present study the test was not speeded, and no answers to this problem can be offered. The data on the speededness are summarized below in tabular form:

	Form 1	Form 2	Form 3
Per cent of examinees who reached the last item..................	97.4	91.8	97.8
Per cent of examinees who reached 98 or more items................	98.9	98.9	98.9

It happened that in each group 264 of the 267 examinees reached Item 98. It is possible that virtually all the unanswered items at the end of the test are in the nature of omitted items, that is, items read but intentionally not answered, rather than items not reached for lack of time.

The inclusion of the twenty special items had a deleterious effect on the test reliability. Reliability coefficients were computed by the Kuder-Richardson formula (20) for the 80-item tests and the 100-item tests. They are presented in Table 3, which also includes the Spearman-Brown predicted coefficients for 100-item tests that would be obtained by adding twenty regular items to the original eighty items.

One might expect the special items to be more or less 'dead wood' in the test, with no major systematic effect on the test reliability. This is essentially the case. It is interesting to note,

TABLE 3.—COEFFICIENTS OF RELIABILITY

Form	80-Item Test (Score A)	100-Item Test (Score D)	100-Item Test Predicted from 80-Item Test	Difference Column (3) Minus Column (4)
(1)	(2)	(3)	(4)	(5)
1.....	.892	.879	.912	−.033
2.....	.896	.892	.915	−.023
3.....	.880	.866	.902	−.036

however, not only that all three coefficients in Column (3) of the table are lower than corresponding values in Column (2) but also that the decrease, small though it may be, is in this instance directly related to the number of responses made to the special items. The differences in Column (5) show the loss in efficiency due to the presence of the special items.

Let us now consider item statistics. The difficulty of each item was measured in terms of 'delta.' Delta is computed by first obtaining the normal deviate which corresponds to the proportion answering the item correctly and then transmuting it to a scale with mean of 13 and standard deviation of 4. The higher the delta, the more difficult the item. The biserial correlation between item score and criterion score was computed twice for each item: once with Score A as criterion and once with Score D as criterion. It was expected that the presence of the special items in the criterion score would decrease the correlations for the regular items. Means and standard deviations of these statistics are presented in Table 4.

The difference of .24 between the mean delta for the regular items of Form 2 and the means for Forms 1 and 3 appears too small to be of great import. In the case of the twenty special items, the differences between the mean deltas are greater than the differences for the regular items. All three differences between forms are statistically significant beyond the one per cent level of confidence, even though only twenty pairs of observations are involved in each instance.[3] For very difficult items a change in directions is likely to result in a change in the index of item difficulty sufficient to affect the precision of item

---

[3] The standard errors of these and other differences based on the data of Table 4 were computed from the formula, $\sigma_d^2 = \sigma_a^2 + \sigma_b^2 - 2\sigma_a\sigma_b r_{ab}$.

The effect of the different instructions on examinee behavior is strikingly portrayed in another fashion in Table 2, which gives frequency distributions of the number of omissions per item. (An unanswered item is considered an omission only if a later item has been answered. Otherwise it is classified as 'not reached.') Similarities and differences are so clear-cut that tests of statistical significance are not necessary. Under all the conditions the special items were omitted more frequently than the regular, easier items, but the difference between special and regular items is particularly great for Form 2 (instructions not to guess). Only a few more omissions occur for Form 1 (no instructions on guessing) than for Form 3 (instructions to answer every item).

If the different instructions affect the typical examinee's rate of work, proper timing under various conditions becomes an important problem. The use of a speeded test would be necessary to establish typical rates of work. In the present study the test was not speeded, and no answers to this problem can be offered. The data on the speededness are summarized below in tabular form:

	Form 1	Form 2	Form 3
Per cent of examinees who reached the last item..................	97.4	91.8	97.8
Per cent of examinees who reached 98 or more items................	98.9	98.9	98.9

It happened that in each group 264 of the 267 examinees reached Item 98. It is possible that virtually all the unanswered items at the end of the test are in the nature of omitted items, that is, items read but intentionally not answered, rather than items not reached for lack of time.

The inclusion of the twenty special items had a deleterious effect on the test reliability. Reliability coefficients were computed by the Kuder-Richardson formula (20) for the 80-item tests and the 100-item tests. They are presented in Table 3, which also includes the Spearman-Brown predicted coefficients for 100-item tests that would be obtained by adding twenty regular items to the original eighty items.

One might expect the special items to be more or less 'dead wood' in the test, with no major systematic effect on the test reliability. This is essentially the case. It is interesting to note,

TABLE 3.—COEFFICIENTS OF RELIABILITY

Form	80-Item Test (Score A)	100-Item Test (Score D)	100-Item Test Predicted from 80-Item Test	Difference Column (3) Minus Column (4)
(1)	(2)	(3)	(4)	(5)
1.....	.892	.879	.912	−.033
2.....	.896	.892	.915	−.023
3.....	.880	.866	.902	−.036

however, not only that all three coefficients in Column (3) of the table are lower than corresponding values in Column (2) but also that the decrease, small though it may be, is in this instance directly related to the number of responses made to the special items. The differences in Column (5) show the loss in efficiency due to the presence of the special items.

Let us now consider item statistics. The difficulty of each item was measured in terms of 'delta.' Delta is computed by first obtaining the normal deviate which corresponds to the proportion answering the item correctly and then transmuting it to a scale with mean of 13 and standard deviation of 4. The higher the delta, the more difficult the item. The biserial correlation between item score and criterion score was computed twice for each item: once with Score A as criterion and once with Score D as criterion. It was expected that the presence of the special items in the criterion score would decrease the correlations for the regular items. Means and standard deviations of these statistics are presented in Table 4.

The difference of .24 between the mean delta for the regular items of Form 2 and the means for Forms 1 and 3 appears too small to be of great import. In the case of the twenty special items, the differences between the mean deltas are greater than the differences for the regular items. All three differences between forms are statistically significant beyond the one per cent level of confidence, even though only twenty pairs of observations are involved in each instance.[3] For very difficult items a change in directions is likely to result in a change in the index of item difficulty sufficient to affect the precision of item

---

[3] The standard errors of these and other differences based on the data of Table 4 were computed from the formula, $\sigma_d^2 = \sigma_a^2 + \sigma_b^2 - 2\sigma_a\sigma_b r_{ab}$.

TABLE 4.—MEANS AND STANDARD DEVIATIONS OF ITEM STATISTICS

Statistic	Form 1		Form 2		Form 3	
	Mean	SD	Mean	SD	Mean	SD
Delta						
80 regular items	12.81	2.58	13.05	2.71	12.81	2.61
20 special items	16.26	1.18	17.63	1.17	16.02	1.10
r_{bis}						
80 regular items:						
Score A criterion	.440	.115	.453	.119	.423	.127
Score D criterion	.433	.110	.446	.113	.417	.128
20 special items:						
Score A criterion	.070	.148	.130	.109	.086	.146
Score D criterion	.127	.138	.220	.109	.140	.152

equating.[4] Without further experimentation, this generalization applies to additional items but does not extend beyond the particular groups employed, although there is no reason to believe that the present groups are atypical with respect to these results. Similar experiments could be expected to yield results of the same nature as those reported here.

All mean biserial correlations for the regular items computed with Score D as criterion are smaller than the corresponding means with Score A as criterion. Although the differences are statistically significant beyond the one per cent level of confidence, they are very small, too small to be of any practical consequence. For the three forms the means of the differences are, respectively, .0067, .0075, and .0067. The standard deviations are .0164, .0243, and .0184; and the standard errors of the means are .0019, .0028, and .0021. For the twenty special items a significant increase appears. This result is without doubt due to the fact that the special-item scores are included in Score D but not in Score A, and it can be explained entirely on this basis.

Differences in mean biserial correlations between forms tend not to be statistically significant. For the regular items the two Form 2-minus-Form 3 differences are significant at the five per cent level but not at the one per cent level. In the case of the special items, none of the differences between values in the first row is significant, but when Score D is used as criterion the Form 2 mean is greater than either of the other means by amounts which are significant at the five per cent level of confidence.

---

[4] For a discussion of item equating see L. L. Thurstone, "The calibration of test items," *The American Psychologist*, II (March, 1947), 103–4.

The findings of this study may be summarized in part in terms of the hypotheses set down at the beginning of this report.

1) There did occur some guessing on the part of all three groups. For no group was the mean number of responses to the set of special items less than fifty per cent. (The special items are the very difficult items and nonsense items, which invite guessing.)

2) The amount of guessing varied with differences in instructions. The variation was slight but statistically significant between the group receiving no instructions about guessing and the group told to answer all the items. The group told not to guess responded to substantially fewer items than either of the other two groups.

3) The relationship between ability and the tendency to guess is very low. The correlation between the score on the regular items and the number of responses to the special items is positive for each group, but only one coefficient, that for the group told not to guess, differs significantly from zero.

Other findings and implications are as follows:

4) The classifications, 'very difficult' and 'nonsense,' may be useful only to the test writer. To the examinee there may be no difference, for he appears to respond in the same way to both. Thus, a very difficult item, though it may have a perfectly sound answer, may act in the test in the same way as though it were a nonsense item.

5) Too many very difficult items in a test can lower its reliability.

6) Instructions which discourage guessing may reduce accurate comparison of groups and test forms when the comparison involves also a test with instructions to guess or no instructions about guessing, where such comparisons are made through measures of item difficulty.

7) Measures of internal validity of individual items are not seriously affected by including some very difficult items in the criterion. In the present study twenty per cent of the criterion Score D were 'special' items. Normally, the percentage of difficult items would be lower than this figure.

RELATIVE CONTRIBUTIONS OF APTITUDE AND WORK HABITS TO ACHIEVEMENT IN COLLEGE MATHEMATICS[1]

WILLIAM C. KRATHWOHL

Institute for Psychological Services
Illinois Institute of Technology

A question which frequently arises is, "How much is achievement in a subject influenced by ability and how much is it influenced by such personality factors as work habits of industriousness and indolence?"[2] A solution to this problem for English was given by Krathwohl (*5,6*). He found that a measure of work habits of industriousness and indolence for English could be secured by defining an index of industriousness for English to be equal to the score, for an individual, on an English achievement test minus his score on a vocabulary test provided both scores were based on the same group (*7*). The scores used for both tests were derived scores which have a mean of 20 and a standard deviation of 4. These scores can readily be transformed to the more familiar standard scores with a mean of 50 and a standard deviation of 10 by multiplying the derived scores by 2½. By means of this index of industriousness for English he found that if students were grouped into above average, average, and below average groups according to ability in English as measured by vocabulary scores, work habits contributed as much or more toward achievement in English as did vocabulary. If, however, students were grouped according to indexes of industriousness for English into industrious, normal and indolent groups, practically all of the variance of achievement in English was accounted for by the vocabulary scores and practically none by the indexes of industriousness for English. Nevertheless, the interesting fact appeared that far superior prediction results for English achievement were obtained from vocabulary scores than by means of any other method, provided the predictions were made separately for each of three groups—the industrious, normal, and indolent groups.

---

[1] Presented at the Annual Meeting of the Midwestern Psychological Association at Cleveland, Ohio, April 25, 1952.

[2] For conciseness and also to avoid awkward construction, the word indolence, as employed in this investigation is used not in a derogatory sense, but rather as a substitute for under-achievement. In the same way, the word industriousness is used as a substitute for over-achievement.

The question now becomes, can the same conclusions be drawn for mathematics as were drawn for English, particularly since it was shown by Krathwohl (*8*) that work habits in mathematics are independent of work habits in English. In other words, he found that an individual could be very industrious in mathematics and at the same time could be very indolent in English. To answer this question it is necessary to measure work habits for mathematics. This was done (*3,4*) by defining an index of industriousness for mathematics to be the score that an individual makes on a mathematics achievement test minus his score on a mathematics aptitude test. The scores used for both tests were derived scores which have a mean of 20 and a standard deviation of 4. For college algebra, the scores used were the grades A, B, C, D, and E which were replaced by the numbers 3, 2, 1, 0, and −1, respectively.

In order to ascertain if the results previously found for English also held for mathematics, a group of 859 freshmen at the Illinois Institute of Technology were selected who took orientation tests between September, 1947, and February, 1949. The mathematics aptitude test selected was the Iowa Mathematics Aptitude Test, Form M. The mathematics achievement test which was used to compute the index of industriousness for mathematics was a locally prepared test on algebra and mensuration, called the Mathematics Preparation Test. Both of these tests were given before the student entered the Institute. The college mathematics achievement test from which the relative effects of aptitude and work habits were to be computed was college algebra, which was given at the end of the first term to students who were properly prepared in algebra and one term later if a review of high-school algebra were necessary. The investigation of the relative contribution of mathematics aptitude and indexes of industriousness for mathematics was made by means of multiple correlation techniques and the results of the investigation are shown in Table I.

In Table I, the columns are numbered from 1 to 8 to facilitate discussion. Wherever the subscript 1 appears, it refers to the college algebra grade, the subscript 2 to the mathematics aptitude

TABLE I.—INTERCORRELATION AND MULTIPLE CORRELATION COEFFICIENTS AMONG (1) COLLEGE ALGEBRA, (2) MATHEMATICS APTITUDE, AND (3) INDEXES OF INDUSTRIOUSNESS FOR MATHEMATICS WITH THEIR RELATIVE CONTRIBUTIONS

Column No.	1	2	3	4	5	6	7	8
Group	N	r_{12}	r_{13}	r_{23}	$r_{1.23}$	Math. Apt.	I.I.	K^2
All	859	.46	.08	−.51	.59	31	3	65
Above average	286	.34	.41	−.04	.53	12	17	71
Average	362	.10	.37	−.27	.42	2	16	82
Below average	211	.17	.21	−.44	.37	5	7	87
Industrious	216	.53	−.08	−.44	.56	33	−1	69
Normal	430	.59	.08	−.18	.62	37	2	62
Indolent	213	.49	.15	−.29	.58	29	5	67

derived score and the subscript 3 to the index of industriousness or I.I. for mathematics.

The first column gives the frequency of the group being investigated. The second column gives the correlation coefficients between college algebra grades and mathematics aptitude scores. The third column gives the correlation coefficients between college algebra grades and indexes of industriousness or I.I.'s for mathematics. The fourth column gives the correlation coefficients between mathematics aptitude scores and I.I.'s for mathematics. The fifth column gives the multiple correlation coefficients for college algebra grades when account is taken both of the mathematics aptitude scores and the I.I.'s for mathematics. The sixth column gives the percentage of variance in college algebra grades which is contributed by the mathematics aptitude test. The seventh column gives the percentage of variance which is contributed by the I.I.'s, and the eighth column gives the percentage of variance still to be accounted for.

It happens in this table that a sharp distinction can be made easily between the correlation coefficients which are significant at the one per cent level and those which are not. All coefficients whose absolute values are greater than 0.17 are significant at better than the one per cent level, whereas, all coefficients equal to or less than 0.17, either are significant at the five per cent level or else are not significant.

In the first row, where the entire group is considered, the second column shows that the correlation between college algebra grades and mathematics aptitude scores is equal to 0.46 and is significant at well above the one per cent level. The percentage of variance in the sixth column, which the mathematics aptitude score contributes to college algebra grades, is thirty-one per cent, the percentage of variance which the I.I.'s contribute is three per cent and the per cent of variance still to be accounted for is sixty-five per cent. The relatively large negative coefficient in column 4 of -0.51 between mathematics aptitude and I.I.'s for mathematics shows the tendency for increased industriousness on the part of students with low mathematical ability and may be due to screening effects during the first term which operated to eliminate the less industrious students. The correlation coefficient in the third column between college algebra and I.I.'s for mathematics of .08 shows that if the group is taken as a whole there is little if any correlation between grades and work habits. This is also shown by the seventh column where the contribution of I.I.'s to grades is only three per cent.

The combination of a smaller correlation between college algebra and mathematics aptitude than is usually found, the moderate contribution of mathematics aptitude to achievement and the rather high unknown variance is partly due to homogeneity of the group, they being required to pass an entrance examination involving mathematics achievement. Other factors which tended to lower the correlation coefficient between mathematics aptitude and college algebra grades were that only five grades were awarded, partly on a subjective basis, and that grades given by college instructors are often not as reliable as they might be (Krathwohl (*2*)).

Previous investigations with indexes of industriousness have sometimes shown that using the entire group conceals some variations. By resolving that group into smaller ones these variations become apparent.

The first division of the entire group in Table I, into above average, average, and below average groups was made on the basis of the derived scores of individuals on the mathematics aptitude test. Scores for the above average group were 23 and above, for the average group from 18 to 22 and for the below average group 17 and below. The theoretical frequency percentages for three groups were twenty-seven per cent, forty-six per cent and twenty-seven per cent, respectively, or approximately the highest quarter, the middle half and the lowest quarter respectively. The divergence of the actual frequency percentages—thirty-three per cent, forty-two per cent and twenty-five per cent—of these three groups from the theoretical frequency percentages may be due to the screening effect of the entrance examinations, because the above average group has more than its share, while the average group has less.

The second division of the entire group into industrious, normal and indolent groups was made on the basis of the indexes of industriousness for mathematics. Indexes of industriousness for the industrious group were 3 and above; for the normal group from -2 to plus 2; and for the indolent group, -3 and below. It can easily be shown that if the aptitude test, the first achievement test from which I.I.'s are computed, and the indexes of industriousness for mathematics form normal frequency distributions, and if derived scores are used, that the standard deviation of the I.I's is equal to $4\sqrt{2(1 - r)}$ where r is the correlation coefficient between the aptitude test and the achievement test from which the I.I.'s are computed. In this case r is equal to 0.52. If it is assumed that the I.I.'s are continuous instead of discrete variables so that the lower limit for the industrious group is 2.5 instead of 3, the theoretical frequency percentages of the industrious, normal and indolent groups are twenty-six per cent, forty-eight per cent and twenty-six per cent, respectively, or roughly the highest quarter, the middle half and the lowest quarter, respectively. Since the actual frequency percentages of the industrious, normal and indolent groups are twenty-five per cent, fifty per cent and twenty-five per cent, respectively, it is seen that these frequencies come very close to the theoretical figures.

When the entire group is divided into above average, average and below average groups on the basis of the mathematics aptitude scores, all the correlations in column 2 between college algebra and mathematics aptitude are lower than the ones for the entire group. In fact, both of the correlations in column 2, rows 3 and 4 of 0.10 and 0.17 not only are low, but also are not reliable at the one per cent level of significance. Such a lowering of coefficients and reliabilities may be due to the increased homogeneity of these three groups with respect to mathematical ability. The interesting fact is that the three correlation coefficients in column 3 between achievement and I.I.'s for mathematics for the division into ability groupings not only are much higher, but all are significant at well above the one per cent level. Evidently when the students are grouped according to ability more accurate predictions can be made from work habits than from ability. Furthermore, a comparison of columns 6 and 7 for the ability groupings shows that the indexes of industriousness or, in other words, work habits contribute more toward achievement than mathematics aptitude. However, column 8 indicates that with the ability groupings much more of the variance is unaccounted for than for the groupings according to work habits or for the group as a whole.

The three correlations in column 2 for the ability groupings show that with such groupings, predictions of grade average on the basis of ability are useless or of little value. It is true that predictions for achievement can be made instead by means of work habits, but the multiple correlation coefficients for the ability groupings are so much higher than any of the zero order correlations that it is preferable to predict for success with ability groupings by using multiple correlation coefficients. In column 3 the highest coefficients of 0.41 and 0.37 between achievement and work habits for the above average and average groups, respectively, shows that it is worth while in a counseling situation to advise students of above average and average scores in mathematics aptitude, of the possibility of securing greater achievement by increased industriousness. In column 4 for the below average group, the coefficient of −0.44, which is of appreciable size, between mathematics aptitude and indexes of industriousness for mathematics, indicates a greater tendency for the below average group to compensate for their lack of ability by means of increased industriousness than for the above average and average groups.

When the original group is divided into industrious, normal and indolent groups on the basis of indexes of industriousness for mathematics, the three correlation coefficients between college algebra grades and mathematics aptitude scores all become, without exception, higher than any of the other coefficients in column 2 and all are significant at well above the one per cent level. Column 3 for these three groups indicates a negligible correlation between achievement and work habits, probably because the effect of work habits has been taken care of in the method of division itself. Column 4 in which all three correlation coefficients between mathematics aptitude and indexes of industriousness for mathematics are negative and significant shows the tendency for the less able students to be more industrious than the others. This fact is particularly true for the industrious group, where the negative coefficient of −0.44 indicates the strongest tendency as compared with the normal and indolent groups for the less able students to compensate for lack of ability by means of increased industriousness. Columns 6 and 7 show that with work habit groupings, mathematics aptitude contributes much more toward success in college algebra than work habits. In fact, the contributions of work habits are practically negligible.

In column 8 the large unaccounted-for variance toward achievement in college algebra for all divisions and for the group as a whole indicates that many other factors are at work toward success besides mathematics aptitude and work habits. Some of these doubtless are reading ability, general intelligence, and methods of awarding grades.

If zero order correlation coefficients are used, the most accurate prediction for achievement in college algebra is made by dividing the original group into industrious, normal and indolent groups, and predicting achievement separately for each group by means of their respective regression equations.

The answer to the question previously proposed as to whether the same conclusions can be drawn for mathematics as were drawn for English, can now be answered in the affirmative. A comparison of this study for mathematics with the previous one for English shows a remarkable similarity between the two in spite of the fact that work habits for mathematics are independent of those for English. The principal difference is that all of the correlation coefficients for English are larger than are those for mathematics. These differences are undoubtedly due to the use of an objective test with twenty-one scores in measuring English achievement, whereas instructors grades with a range of five scores were used to measure achievement in mathematics.

The similarities which were observed between mathematics and English are the following:

1) When the group is considered as a whole, predictions for achievement are independent of work habits and the contributions of indexes of industriousness both for mathematics and for English are negligible. Also, individuals with lower aptitudes tend to have higher indexes of industriousness.

2) The multiple correlation coefficient for achievement based on aptitudes is higher by about 0.12 for both English and mathematics, than when zero order correlation coefficients are used with the respective aptitudes.

3) When the original group is divided, using aptitude groupings, into above average, average and below average groups the following similarities are noticed: (a) Correlation coefficients between achievement and aptitude become lower than if the single original group is used and in some cases become so low as to be not satisfactorily significant. (b) Correlation coefficients between achievement and work habits, with one exception, are higher for each group for both subjects than are those between achievement and aptitude, and all of these correlation coefficients are significant at the one per cent level.

4) When the original group is divided into industrious, normal and indolent groups on the basis of work habits, the following similarities are observed: (a) All correlation coefficients between achievement and ability are higher than are those for ability groupings and in nearly all cases are higher than when the group is considered as a whole. (b) Correlations between achievement and work habits are small and are not statistically significant. (c) Correlations between aptitudes and work habits on both subjects are, with one exception, negative, indicating that low ability students have to work harder and do work harder than their brighter companions. (d) There is a tendency for the industrious and normal students to have higher correlation coefficients between ability and achievement than for the indolent students. This characteristic is similar to that found by Hartshorne and May in their studies of the social habits of honesty, truthfulness and morality (*1*). They found that the possession

of desirable traits is accompanied by increased consistency and predictiveness.

5) With both mathematics and English the most accurate predictions for achievement are made on the basis of aptitudes when the groups are divided into industrious, normal and indolent groups on the basis of work habits, provided the predictions are made separately for the industrious, normal and indolent groups.

6) The general conclusion which can be drawn from this study is that in spite of the fact that work habits of industriousness for English and for mathematics are independent of each other, their effect on achievement in their respective subjects is very similar.

REFERENCES

1) H. Hartshorne and Mark A. May. *Studies in the Organization of Character*. New York: Macmillan Company, 1930.

2) William C. Krathwohl. "Prediction of average class achievement by means of aptitude tests," *Journal of Engineering Education*, November, 1946, 37, 234–242.

3) ———. "Persistence in college of industrious and indolent work habits," *Journal of Educational Research*, January, 1949, 42, 365–370.

4) ———. "Effects of industrious and indolent work habits on grade prediction in college mathematics," *Journal of Educational Research*, September, 1949, 43, 32–40.

5) "An index of industriousness for English," *Journal of Educational Psychology*, December, 1949, 40, 469–481.

6) ———. "Relative contributions of vocabulary and an index of industriousness for English to achievement in English," *Journal of Educational Psychology*, February, 1951, 42, 97–104.

7) ———. "A theory of work habits of industriousness," *Journal of Engineering Education*, November, 1951, 42, 157–163.

8) ———. "Specificity of over- and under-achievement in college courses," *Journal of Applied Psychology*. April, 1952, 36, 103–106.

READING FOR PROBLEM-SOLVING IN SCIENCE[1]

J. HARLAN SHORES and J. L. SAUPE

University of Illinois

The measurement of reading rate and comprehension has been the subject of numerous researches since the development of objective measuring devices. Few areas have received as large a share of the interest and energies of test-makers. Many of these researches were attempts to discover the basic nature of the reading process and subsequently to improve the measurement of this process. The reading tests so constructed and now in use are evidently based on the existence of a general ability to read (*5*). In general they include such sections as vocabulary, ability to follow directions, ability to grasp the central thought of a passage, and general comprehension (word, phrase, sentence and paragraph). In grades four, five and six, typical items measuring comprehension require the ability to grasp and retain facts contained in a short paragraph.

Within recent years doubt has been cast concerning the existence or at least the value of the concept of a generalized ability to read beyond the primary grades (*2,3,13,14,15,17*). Rather it has been hypothesized that reading skills differentiate with many variable factors, each of which, when varied from one test situation to another, would affect a student's test score. One analysis lists fifteen such factors which should be recognized and either held constant or measured (*14*). Three factors mentioned as having received very little consideration in reading test construction in the past are content area of material being read, reader's purpose (what he intends to get from the reading) and reader's experience background for the specific content of the reading passage (*14*).

Recent investigations leave little doubt that reading rate and comprehension are affected by the kind of material being read. It seems reasonable to expect that the reading skills required for science material will differ from those required for materials of history, mathematics, or other content areas, each of which requires its peculiar combination of abilities. Certainly good

---

[1] This study was made possible by a research grant from the College of Education, Bureau of Research and Service, University of Illinois. The research design and conclusions, however, are those of the authors.

readers in one content area can be expected to read well in other areas and poor readers in one area can be expected to read poorly in other areas, but there are differences within these groups which cannot be explained by the concept of a general ability. Tinker (*19*) and McCallister (*9*) support the contention that reading ability differentiates with the content field in which the reading is done.

The results of these and other investigations which point to the denial of a general ability to read regardless of the kind of material being read, imply that for reading test scores to be of real value they should be reported in terms of ability to read in different content areas. There would need be separate tests of ability to read historical materials, scientific materials, and the like. Only in this manner would it be possible to determine an individual's variable proficiencies in the various content areas.

Experimental evidence and expert opinion each support the theory that the specific purpose in the reader's mind when he approaches a work-type reading task is a major determinant of both reading rate and comprehension. More than twenty years ago Gray pointed to the effect of this factor on reading rate and comprehension (*7*). Since that time the results of research and considered judgment of scholars in the field of reading have supported him by reporting a relationship between purpose and reading rate and comprehension (*1,4,11,19*). Reading comprehension includes the ability to adjust the rate of reading and the specific skills employed to the purpose for which the material is being read. For test-makers this fact implies that the factor of reader's purpose should receive attention in the construction of reading tests. At present this factor is neglected and neither the test-maker nor the interpreter of test scores can know how proficient the readers might have been if they had been reading for a well defined purpose. Test taking is a special instance of a learning situation. Since learning involves goal seeking and the reader's purpose sets his immediate goals, the test-maker cannot know what he has measured until he can make fairly valid assumptions with respect to the similarity of purpose among the testees. Shores (*14*) suggested that prior to the printed passage of the test the reader should be given a clear purpose for his reading. Similarly, Dolch (*6*) recommended that since modern textbooks are written with a purpose in mind, reading tests should also have a purpose. Ideally reading tests should be as similar as possible in purpose to the purposes for which reading is taught.

With respect to experience background of the reader the test-maker must assume that the readers have had equivalent experience with the specialized subject matter of the reading passage in order for the test to be valid in comparing individuals or groups. A child with a wealth of experience in aviation will comprehend a passage about airplanes better and more rapidly than one who has not had this experience. This requirement of equivalent experience is not easily met and has been violated frequently in reading test construction. Methods for meeting this requirement would be to select the subject of the reading passage in such a manner that it might be assumed that the testees have had little if any specific background for it or to use a great number of passages with the expectation that the effects of experience background would cancel out. At the same time the test passage should be typical of the kind of content and purpose for which measurement is desired.

The failure of current reading tests to take these factors into consideration would prompt the construction of a reading test or battery of tests which account for them more adequately. This work would culminate with a standardized battery of tests containing at least one test employing the content of science, one using the social studies, one arithmetic, and so on. Each test should have a clearly defined purpose stated for the reader at the beginning of each reading passage. Every possible attempt should be made in selecting the subject of the written material to hold the factor of experiential background constant.

For at least the past three decades, theory of method in elementary schools has assumed problem-solving as a primary approach. Thus the growth in the training and use of reading to solve individual and group problems that is evidenced in schools today may be expected to continue. It follows that if the new reading tests are to be of maximum use in predicting success or in measuring the relative position of an individual or a class with respect to the kinds of skills needed for normal classroom activity, the individual tests should be measuring reading as a tool in problem-solving. In other words, what the individual tests would measure is the ability to do that kind of reading ordinarily done in elementary-school classrooms with the materials of science, the social studies, and the like.

As a beginning in the construction of such a test battery there is now in the process of refinement a test at the fourth-, fifth-, and sixth-grade levels employing the content of elementary-school science. It has been called a Test of Reading for Problem-solving in Science. When scores on this test are reported and analyzed the question naturally arises: What are the relationships between that which is measured by this test of reading for problem-solving in science and that measured by other reading tests and tests of general ability commonly used in the public schools?

TESTS EMPLOYED

The Test of Reading for Problem-solving in Science consists of two written passages of approximately eight hundred words each. In selecting the content for each passage, material was chosen to be typical of elementary-school science classes and yet such that children probably would not have come into contact with this particular content. This was an attempt on the part of the test-makers to hold the factor of experience background somewhat constant.

The student is told in the test directions and again immediately prior to the reading of each passage, the purpose for which he is doing the reading, i.e., the problem he is trying to solve. The problem of the first selection is, "What is the Best Way for the Farmer to Keep Grub Worms from Harming His Crops?" The student is told that he is reading the second passage to find out, "Do Plants or Animals Like Those on the Earth Live on Mars?" Following each passage are twenty four-choice multiple-choice type items based on the content of the passage. In general each of the first nineteen items following each passage requires the testee to make inferences from the facts in the passage. Each inference is considered to have some relationship to the desired solution of the problem. The stem of the final item of each part is a statement of the problem the student has been asked to solve. There were also four choices for these final items, and the responses to them were included in the total test score without weighting. The correct alternative to each of these items is the solution of the respective problem which follows most logically from the passage.

Although reading rate is important within broad limits in the classroom situation due to the limited length of the school day, an attempt was made to remove the rate factor by allowing sufficient time for each student to complete the entire test. For the purposes of this study the test scores of the very few students (fewer than three per cent) who did not finish the test were considered not representative of their ability and hence were not used.

All test items showed positive discrimination with the upper and lower fourths of the sample as judged by the total score. The reliability coefficient as computed by the Kuder-Richardson formula was +.82. The nature of the test makes an estimate of statistical validity impossible because it was designed to measure an ability not heretofore measured. Hence logical validity is the necessary approach. For a statement concerning the logical validity of the test see Husbands and Shores (*8*) who report a study in which this test was used.

Tests of achievement and mental ability were administered to provide data for the relationships to be described later. These are as follows:

1) New California Short-form Test of Mental Maturity, Primary or Elementary Battery (the appropriate form was administered at each grade level) '47 S-form.

2) Progressive Achievement Tests, Primary or Elementary (the appropriate form was administered at each grade level) Battery, Form A.

In addition, sociometric measures of acceptance and rejection were taken. However, the correlations of these measures with all of the other test scores were so low and generally inconclusive that they are not included in this report.

A complete list of the scores used for each of the 182 cases is:

1) Test score, Reading for Problem-solving in Science (referred to in the tables as Science Reading)
2) Mental Age, language
3) Mental Age, non-language
4) Reading Age
5) Arithmetic Age
6) Chronological Age

METHOD OF INVESTIGATION

This study was conducted in a city of approximately eight thousand population in central Illinois. Classes were chosen in schools located in the middle socio-economic categories. All the pupils in each classroom were tested.

The test of Reading for Problem-solving in Science was administered to two hundred fourteen fourth-, fifth-, and sixth-grade children. Of the two hundred fourteen cases for which scores were taken on the reading for problem-solving test there were one hundred eighty-two cases for which there was complete data on the California Tests of Mental Maturity and the Progressive Achievement Tests.[2]

TABLE I.—INTERRELATIONSHIPS BETWEEN READING FOR PROBLEM-SOLVING IN SCIENCE AND OTHER-MEASURED ABILITIES*,**

	Science Reading	California Language M. A.	California Non-Language M. A.	Progressive Reading Age	Progressive Arithmetic Age	C. A.	N = 182	
							Mean	S.D.
Science Reading	.82	.61	.49	.63	.59	.08	23.75	6.47
California Language M. A.	.61	.95	.53	.81	.73	.29	124.50	20.39
California Non-Language M. A.	.49	.53	.91	.60	.64	.33	128.29	23.63
Progressive Reading Age	.63	.81	.60	.90	.83	.35	129.11	16.71
Progressive Arithmetic Age	.59	.73	.64	.83	.93	.44	128.53	12.24
C. A.	.08	.29	.33	.35	.44	1.00	123.06	14.61

* Product moment correlation coefficients uncorrected.

** Self-correlations are reliability estimates.

---

[2] This discrepancy is due largely to absences when the various tests were administered and children moving into and away from the school district during the period when data were collected.

It is important to note that the data from all the tests except the Test of Reading for Problem-solving in Science were collected from the same students one full year previous to this investigation in connection with another study.[3] Consequently, any marked irregularities in rate of growth of these attributes during the intervening year would disturb the results reported.

TABLE 2.—CORRECTED INTERRELATIONSHIPS BETWEEN READING FOR PROBLEM-SOLVING IN SCIENCE AND OTHER MEASURED ABILITIES*

	Science Reading	M. A. Language	M. A. Non-Language	Reading Age	Arithmetic Age	C. A.
Science Reading		.69	.57	.73	.68	.09
M. A. Language	.69		.58	.88	.78	.30
M. A. Non-Language	.57	.58		.66	.70	.34
Reading Age	.73	.88	.66		.91	.37
Arithmetic Age	.68	.78	.70	.91		.46
C. A.	.09	.30	.34	.37	.46	

* Product moment correlation coefficients corrected for attenuation.

The data reported in Table 1 are product moment correlation coefficients computed from the raw scores. These data do not account for differences possibly due to the relative reliability of the measuring instruments. The data in Table 2 are corrected for attenuation and provide a better estimate of the true relationships among these factors.[4,5]

---

[3] G. Orville Johnson. "A study of the social position of mentally handicapped children in the regular grades." *American Journal of Mental Deficiency*, 55, No. 1, July, 1950.

[4] The formula employed to correct for attenuation is $r_{\infty\infty} = \frac{r_{xy}}{\sqrt{r_{x_1x_2}}\sqrt{r_{y_1y_2}}}$, where $r_{\infty\infty}$ is the corrected correlation coefficient, r_{xy} is the uncorrected product moment correlation coefficient, and $r_{x_1x_2}$ and $r_{y_1y_2}$ are the respective reliability estimates of the two instruments whose scores are being correlated.

[5] Quinn McNemar. *Psychological Statistics.* New York: John Wiley and Sons, 1949, p. 134.

RESULTS

Limitations of sampling and the fact that the Test of Reading for Problem-solving in Science is still being revised suggest that this type of study should be repeated at a later date when more adequate instruments are available. At this stage factor analysis would seem to be appropriate to discover what the relationships are among the Reading for Problem-solving in Science test, current reading tests, and other tests of various mental abilities. The generalizations from the present study should be regarded as tentative. While many of the correlation coefficients reported in Table 2 are not statistically different from one another and others are not significantly different from zero,[6] there may be value in pointing to some relationships which seem to exist between reading for problem-solving in science and the other measured factors.

1) Intercorrelations among the first five measures listed in Tables 1 and 2 (Reading for Problem-solving in Science, Mental Age—Language, Mental Age—Non-Language, Reading Age, and Arithmetic Age) are significantly positive in each instance. This indicates some general ability measured in common by these tests and which is also present in the Test of Reading for Problem-solving in Science.

2) Reading to solve problems in science correlates highest with reading age and higher with language mental age than with non-language mental age. A major factor in this type of reading is probably the reading-language factor.

3) Reading for problem-solving in science correlates lower with each of the other of the first five factors listed in Table 2 than does language mental age and reading age. The important implication here is that this test is measuring an ability which has less of the general factor causing high intercorrelations among all of them. The suggestion is that ability to do the type of work-type reading required by problems in science, a reading skill which involves both reading and thinking critically about that which is read, is more independent of mental age than is general reading ability and is different in some degree from whatever is measured in tests of general verbal intelligence and general ability to read.

---

[6] Correlations within the range of +.18 to −.18 are not regarded as significantly different from zero at the one per cent level of probability.

4) Ability to read to solve problems in science correlates significantly lower with chronological age than do any of the other measures of mental ability or achievement. This suggests that this ability is nurtured less by maturation and incidental cultural impact than are the other measured abilities. It also suggests that significant development of this ability probably requires deliberately planned learning situations not uniformly provided in the schools employed in this experiment.

5) Ability to read in order to solve problems in science correlates lower with Reading Age and Language Mental Age than do these two abilities with one another. Again the indication is that this ability is somewhat dissimilar to that measured as verbal intelligence or general ability to read.

SUMMARY

Considerable evidence is accumulating to support the hypothesis that reading ability differentiates beyond the primary grades into somewhat specific abilities to read different kinds of material for different purposes. Research along these lines continues to be hampered by the lack of adequate instruments for measuring whatever form these differentiated abilities assume. This investigation, using an instrument which after considerable development is still being revised, tends to support the hypothesis that reading of the kind employed in grades four, five and six to solve problems in science has a large factor in common with mental ability and general achievement as these are commonly measured and yet is somewhat unique in a manner which cannot be accounted for by these generalized factors. A reasonable prediction is that sharper measuring instruments will not only substantiate the hypothesis that general ability to read does differentiate into specific abilities, but will also describe the extent and nature of this differentiation and the amount and character of the remaining common general factors.

BIBLIOGRAPHY

1) P. J. Bloomers. "Rate of comprehension of reading: Its measurement and its relationship to comprehension." *Journal Educational Psychol.*, xxxv, November, 1944, pp. 449–472.

2) E. A. Bond. *Tenth-grade Abilities and Achievements.* New York: Bureau of Publications, Teachers College, Columbia University, 1938.

3) E. Bond. *Reading and Ninth-grade Achievement.* New York: Bureau of Publications, Teachers College, Columbia University, 1938.

4) H. M. Carpenter and M. A. Young. "Reading to Learn History." In National Council for the Social Studies, Seventeenth Yearbook, *Study and Teaching of American History.* Washington, D. C.: The American Council on Education, 1949, pp. 285–304.

5) F. B. Davis. "What do reading tests really measure?" *English Journal*, XXXIII, April, 1944, pp. 180–187.

6) E. W. Dolch, "Teaching reading." *Elementary School Journal*, XXXIV, September, 1933, pp. 36–43.

7) W. S. Gray. "The relation between study and reading." *Addresses and Proceedings.* Washington, D. C.: National Education Association, 1919, pp. 580–586.

8) K. L. Husbands and J. H. Shores. "Are fast readers the best readers." *Elementary English*, XXVII, January, 1950, pp. 52–57.

9) J. M. McCallister. "Reading difficulties in studying content subjects." *Elementary School Journal*, XXXI, November, 1930, pp. 191–201.

10) C. M. McCullough and others. *Problems in the Improvement of Reading.* New York: McGraw-Hill Book Co., 1946.

11) A. McIntire. "Reading social studies materials in the middle grades." *Elementary English Review*, XXI, November, 1944, pp. 262–266.

12) Q. McNemar. *Psychol. Statistics.* New York: John Wiley and Sons, Inc., 1949.

13) J. H. Shores. *The Ability to Read Historical Materials as Related to Eighth-grade Achievement and General Reading Ability.* Unpublished master's thesis. University of Minnesota, 1938.

14) J. H. Shores. "Some considerations of invalidities of general reading tests." *Journal of Educational Res.*, XL, February, 1947, pp. 448–457.

15) J. H. Shores. "Skills related to the ability to read history and science." *Journal of Educational Res.*, XXXVI, April, 1943, pp. 584–593.

16) E. T. Sullivan, W. W. Clark, and E. W. Tiegs. *New California Short-form Test of Mental Maturity—'47 S-form*, Los Angeles: California Test Bureau, 1947.

17) E. J. Swenson. "A study of the relationships among various types of reading scores on general and science materials." *Journal of Educational Res.*, XXXVI, October, 1942, pp. 81–90.

18) E. W. Tiegs and W. W. Clark. *Progressive Achievement Tests.* Los Angeles: California Test Bureau, 1943.

19) M. A. Tinker. "Speed vs. comprehension in reading as affected by level of difficulty." *Journal of Educational Psychol.*, XXXVIII, April, 1947, pp. 389–396.

INTERRELATIONSHIPS AMONG MATERIALS READ, WRITTEN AND SPOKEN BY PUPILS OF THE FIFTH AND SIXTH GRADES

HENRY R. FEA

Sacramento State College, Sacramento, California

The close relationships among the four language arts—speaking, listening, reading and writing—have been explored by research workers and advocated by curriculum-makers. A more intensive study of interrelationships among speaking, reading, and writing was attempted in the present study by applying nine measures to materials read, written and spoken on the same topic by a group of fifth- and sixth-grade pupils. The study employed measures which had previously been used for reading, oral language or written language and applied them simultaneously to samples of all three forms of communication. This made possible comparisons of the different language arts abilities at these grade levels and an analysis of the measures themselves.

More specifically, the study attempted to answer the following questions:

1) What is the level of development in each of the three language arts for the same children in grades five and six?

2) Does varying the oral-written order of reproduction affect the quality of oral or written samples of pupils' work?

3) Does the developmental level of oral and written samples vary more with level of material read than with reading ability of the pupil?

4) Is level of development revealed by one measure comparable with that revealed by others?

5) Are any measures suitable as multiple-measures of the different language arts?

6) If measures prove suitable as multiple-measures, what is the order of development in each factor considered?

PREVIOUS STUDIES

There have been few studies which attempted to measure development in more than one of the language arts simultaneously. Lorge and Kruglov(*18*) investigated the relationship between intelligence and readability level of written compositions, for eighth- and ninth-grade pupils. They noted that readability level of compositions was approximately two grades below expected reading status.

Bushnell(*2*) attempted an analytical contrast of some factors of the oral and written English of tenth-grade pupils. Each pupil gave a short narrative oral composition. Later, the pupils wrote themes on the same subject matter. Bushnell found only nine cases in which oral compositions were superior as measured by the opinion of judges using the Van Wagenen Composition Scale. Errors in sentence structure were more numerous in oral compositions. There were on the average three overloaded or disjointed sentences in every oral theme, but only one such sentence in every eight written themes. Bushnell gives his opinion that oral language is less subject to training in the schools and remains on an immature level as judged by number of words, number of sentences and number of words per sentence as well as in general quality. He admits that sentence length may not be a valid measure because of its dependence on punctuation.

Schonell(*25*) found more cases of backwardness in written language than in spelling or reading. He suggested environment as a reason, stating that reading and spelling are more dependent on direct teaching. He believes that reading affects vocabulary in oral and written language by unconscious assimilation. This does not operate with the same potency on the subtler characteristics of sentence structure. Thus, style and structure do not transfer to the same extent as vocabulary.

Dow and Papp(*5*) investigated relationships among test scores of reading ability, language ability and grades in fundamentals of speech, public speaking and literary interpretation. Their subjects were students in sophomore English courses. Reading scores and scholastic aptitude scores were determined from tests given in freshmen year. Scores on fundamentals of speech, public speaking and literary interpretation were taken from grade books of instructors. They admit that, in light of measures used, validity of their findings may be open to question, but conclude that no significant relationships appeared among reading ability, language ability and speech ability.

Lemon and Buswell(*16*) investigated errors in oral and written expression of twenty ninth-grade children. Oral samples were recordings of informal conversation by a concealed microphone. Written samples were themes obtained as class assignments. For comparison they equalized the two samples from each pupil to the sample containing the smaller number of words. Since the coefficient of correlation between oral and written errors was −.289 they questioned methods of teaching which they felt yielded no transfer in a situation conducive to transfer.

Watts(*29*) studied oral and written language development in children. He contends that success in written language is dependent upon success already having been achieved in oral language. He believes that children will not use the prearranged form characteristic of good prose but use the style of everyday speech which has no prearranged form. Thus he infers that speech is a groping process of clarification of thought, and that this is not true of written language. This point may be open to question. Careful writers 'get something down,' then rewrite for good form.

Mathews, Larsen and Gibbon(*19*) investigated the importance of reading ability for freshmen taking rhetoric classes. They performed three experiments involving teaching composition purely by reading materials. Scholastic aptitude, reading and rhetoric levels showed all of the high grade group to be in the upper quarter in reading skill. The low grade group was only slightly above average in reading skill. All experimental groups showed appreciable improvement in reading and no appreciable lag in grammar as compared to the control group who were given direct instruction in rhetoric.

Rossignol(*21*) explored relationships among hearing acuity, speech production and reading performance of primary-grade children. Hearing acuity was tested by a pure-tone audiometer, speech production by two examiners using an articulation test and a sound-repetitions test. Reading performance was checked by the Gates Primary Reading Test. She found a small but significant relationship between reading performance and speech production.

THE MEASURES

The experimental material consisted of three samples of language from each pupil:

1) A transcription of the oral reading by each pupil of the story

Golden Harvest by Elizabeth Yates(*23*). This story had been adapted previously by use of the Lorge Formula(*17*) to an exact 4.5 reading level.

2) A transcription of the oral reproduction of the same story by each pupil.

3) A transcription of the written reproduction of the same story by each pupil.

The samples were then analyzed by the following measures:

1) Vocabulary—number of words.

2) Vocabulary—number of different words.

3) Vocabulary—number of words not found on the Dale list(*3*).

4) The Type-token Ratio—This measure was used by Johnson(*14*). It is an expression of the ratio of the number of different words (types) to the total number of words (tokens).

5) The Lorge formula for readability(*17*). This is a measure of the reading difficulty of materials written for children, using the number of words, number of difficult words, number of sentences and number of prepositional phrases.

6) The mean and standard deviation of sentence length. This measure was used by Schonell(*24*).

7) Degree of subordination. This measure was used by La Brant(*15*). It is expressed as a ratio of the number of dependent clauses to the number of independent clauses.

8) Number of prepositional phrases. This measure has been used widely; one example is that of Watts(*29*).

9) Some measure of ideas expressed.

THE SUBJECTS

The one hundred forty cases were selected from children of the fifth and sixth grades of four elementary schools in two California cities. Basis for selection of subjects was:

1) Reading ability of grade three or better as revealed by results of the Van Wagenen Unit Scales of Attainment.

2) All children were of the white race and came from homes where English was the only language spoken.

3) Normality of sight, speech and hearing as revealed by school records.

Since it was possible to vary the order of reproduction of oral and written samples, and because variation of such order might affect the quality of the samples, two groups were used. Group

A followed the order; reading, telling, writing: Group B followed the order; reading, writing, telling. The groups were equated on a basis of sex, grade, age, reading grade scores, and socio-economic background.

CONDUCT OF THE EXPERIMENT

Each subject was taken to a small room in his school containing an Audograph recorder. The pupil was informed of the three tasks expected of him, shown the operation of the machine, and told that his reading and telling of the story would be recorded. For recording, the child remained seated, and a lapel microphone was used. Each child was given two minutes to organize his thoughts before telling the story he had just read. To write the story, the pupil was taken to a second small room containing several desks.

When all samples had been procured they were subject to the following transcription and analysis:

1) Oral reading recordings—analyzed according to the Gray Oral Reading Analysis(*8*) for fluency (time required to read the story), mispronunciations, omissions, substitutions, insertions, repetitions, reversals, and faulty phrasing (excessive pausing where no pause is indicated in the text, as used by Hahn(*9*)).

2) Oral reproduction recordings—analyzed for repetitions, unintelligible remarks, punctuation, number of words, number of different words, number of words not appearing on the Dale list(*3*), number of prepositional phrases, number of sentences, number of run-on sentences, number of incomplete sentences (as defined by Hoppes(*12*)), degree of subordination, and number of correct verbal memories (this is a measure of the number of reproduced facts).

3) Written reproductions—analyzed for the same factors as oral reproductions with exception of repetitions and unintelligible remarks.

ANALYSIS OF THE RESULTS

An analysis of the original story as read by the children was made for comparison with their later oral and written samples. Results of this analysis are given in Table I.

The second step was statistical analysis of the children's language samples. Are the measures appropriate as measures of material read, spoken and written? Does application of these measures show relationship among the language samples? Two statistical measures are used in an attempt to answer the second question: correlation, to reveal the degree of relationship; level of significance of difference between means when a measure is applied to two of the media. Obtained correlation coefficients are listed in Table II. Means, standard deviations and critical ratios between means are illustrated in Table III.

TABLE I.—RESULTS OF MEASURES APPLIED TO THE STORY, *Golden Harvest*

Measures	Results
Total number of words	750.
Total number of different words	327.
Total number of hard words	70.
Total number of phrases	65.
Total number of sentences	63.
Degree of subordination	.39
Total number of facts	109.
Average sentence length	11.90
SD of sentence length	5.97
Type-token ratio	.44
Lorge grade rating	4.50

From these results some twenty conclusions may be advanced about the interrelationships of the three language arts and their measures:

1) Relationship among the number of minutes to read, tell and write the story. This measure is misleading because a pupil who speaks or writes quickly with occasional long pauses receives the same score as one who speaks or writes slowly with no long pauses. Table II shows negligible correlation except for oral and written reproduction correlation coefficient of .5. Investigations have shown positive correlation of reading comprehension and speed. Therefore, pupils who read quickly should have greater degree of comprehension and remember more. Because they remember more they should have more to tell and write. Thus, reading time should correlate negatively with oral and written reproduction time. This reasoning disregards the possibility of a general verbal fluence factor which would tend to produce positive

TABLE II.—CORRELATION COEFFICIENTS OF RELATIONSHIPS OF LANGUAGE ARTS FACTORS

Factors	r
Total number of verbal memories: oral with written	.82
Total number of different words: oral with written	.67
Total number of hard words: oral with written	.64
Total number of phrases: oral with written	.63
Total number of words: oral with written	.62
Total number of minutes for reproduction: oral with written	.50
Total number of sentences: oral with written	.43
Type-token ratio: oral with written	.42
Reading grade with number of written verbal memories	.40
Reading grade with number of oral verbal memories	.37
Number of repetitions in reading with number of repetitions in oral reproduction	.33
Excessive phrasing in reading with excessive phrasing in oral reproduction	.32
Average sentence length: oral with written	.17
Lorge rating: oral with written	.16
Reading grade with Lorge rating of oral reproductions	.11
Number of minutes for reading with number of minutes for oral reproduction	.07
Number of minutes for reading with number of minutes for written reproduction	.06
Number of mispronunciations in reading with Lorge rating of written reproductions	−.002
Reading grade with Lorge rating of written reproductions	−.004
Number of mispronunciations in reading with number of oral verbal memories	−.03
Number of mispronunciations in reading with number of written oral memories	−.11
Number of mispronunciations in reading with Lorge rating of oral reproductions	−.11

time relationships. Perhaps both factors operate to cancel the effect of either. Mean reading speed is approximately one hundred seventeen words-per-minute; oral reproduction speed, one hundred eight words-per-minute; written reproduction, eight words-per-minute. The last figure is low in relation to speed of handwriting usually quoted for these grade levels.

2) Relationship of reading grade to number of oral and written verbal memories.—The positive relationship shown in Table II would appear normal. Most reading tests make considerable demand upon immediate memory. It would seem natural that a pupil receiving a high oral score would receive a high written

Table III.—Means, Standard Deviations and Critical Ratios between the Means of Factors in Oral Reproductions and Written Reproductions of One Hundred Forty Pupils

	Oral Samples		Written Samples		Critical Ratio
	Mean	SD	Mean	SD	
Total number of words	291.21	124.15	196.48	81.36	7.55
Number of hard words	18.26	7.96	13.38	6.31	5.67
Number of different words	115.39	39.16	94.16	31.02	5.03
Number of sentences	15.21	6.91	11.35	6.02	4.99
Number of phrases	19.77	10.09	15.47	7.82	3.98
Number of run-on sentences	1.96	1.82	1.33	1.38	3.26
Number of verbal memories	35.01	15.45	30.38	13.60	2.66
Degree of subordination	.32	.09	.32	.10	.01
Average sentence length in words	19.99	5.86	22.85	23.01	−1.43
Number of incomplete sentences	.29	.62	.43	.82	−1.57
Lorge rating	4.50	.37	4.81	1.41	−2.53
Type-token ratio	.43	.08	.50	.08	−8.17

score. The critical ratio of 2.66 is of such magnitude that separate norms would be necessary if the measure were used for both oral and written language.

3) Relationship of the number of reading mispronunciations to the number of oral and written memories.—Table II shows negligible relationship. Probably two opposing factors produce the result. If a pupil cannot pronounce a word and does not know its meaning he will not use it orally or in writing; if he knows the

meaning but has difficulty with the pronunciation, extra time spent trying to pronounce it should fix it in his memory.

4) Relationship of the number of reading mispronunciations to Lorge rating of oral and written samples.—The Lorge formula seems excellent as used with oral samples. Its use in written language samples of children who show no evidence of sentencing is questionable.

5) Relationship of reading grade to Lorge rating of oral and written reproductions.—Results, shown in Table II, indicate little relationship between pupil reading-level and level of difficulty of style of his oral and written samples. However, they may indicate that level of material read has more influence on level of maturity of reproductions than reading level of the pupil.

6) Relationship of excessive phrasing in reading to excessive phrasing in oral reproduction.—Results support the findings of Hahn(*9*) that it is a habit.

7) Relationship of the number of repetitions in reading to the number of repetitions in oral language.—Table II shows a definite tendency for those who repeat in reading to do the same in oral language reproduction.

8) Relationship of the total number of words used in oral and written language situations.—Results of Table II indicate a definite tendency for those who use more words orally to do the same in writing. In only seven cases did a pupil write more words than he spoke. The critical ratio is of such magnitude as to require different norms for the two media but the measure appears to be valid.

9) Relationship of the number of different words used in oral and in written language samples.—The relationship here is greater than for total number of words. But, again, separate norms would be necessary. Pupils in this situation used in speaking approximately thirty-five per cent, and in writing approximately twenty-nine per cent of the number of different words encountered in reading.

10) Relationship of the number of hard words in oral and written language.—Results are similar to those of the two previous measures so this measure is probably superfluous in measuring similarities when the other two are used.

11) Relationship of the number of phrases in oral and written language.—Studies in oral language have been limited to con-

sideration of the age at which phrases appear in the language of the child. Stormzand and O'Shea(*27*) used the measure in written language. On the basis of their findings the present study should have shown the mean number of phrases in oral and written language as approximately thirty-five and twenty-four. Actual results as shown in Table III are 19.77 and 15.47 phrases per sample. Perhaps difference in material would account for this discrepancy as Stormzand and O'Shea used original material. Table II shows a definite relationship for this measure; therefore, it is valid but separate norms would be necessary.

12) Relationship of the number of sentences in oral and written language.—Placing responsibility with the pupil for indication of sentencing has been used here. Previous investigators have considered the thought-unit as synonymous with a sentence. This makes the sentence count subjective and it has been condemned by Watts(*29*), Johnson(*14*), Seegers(*26*) and La Brant(*15*) for this reason. In this study the sentence-unit appears sufficiently valid in oral language to justify its use. But due to inability of a few pupils to use punctuation, results obtained from written samples are questionable. Results are at variance with those of Bushnell(*2*) and Bear(*1*) in that the present study shows more sentencing.

13) Relationship of the number of run-on sentences in oral and written language.—Results shown in Table III substantiate the findings of Wiswall(*30*). The measure is vulnerable to the extent that the definition of a sentence is subjective.

14) Relationship of the number of incomplete sentences in oral and written language.—This measure tends to be more objective than a sentence count. Table III shows this to be the first measure considered where the mean for written samples exceeds that for oral samples. Also, this measure could be applied in both media using the same norms. However, it is an unsuitable measure at these grade levels as only thirty-one oral and thirty-nine written samples contained incomplete sentences.

15) Relationship of the degree of subordination in oral and written language.—This measure has been used by many previous investigators such as Heider and Heider(*10*) and La Brant(*15*). From evidence of Tables II and III this would appear to be a suitable measure of oral and written language samples. The same norms would be suitable for both media.

16) Relationship of the number of correct verbal memories in oral and written language.—There is high degree of relationship here.

17) Relationship of the average sentence length in oral and written language.—There is, according to Table II, low degree of relationship. The questionable validity of the sentence as a measure has been previously discussed.

18) Relationship of the standard deviation of sentence length in oral and written language.—Result for oral language is similar to the original story: result for written language is very largely due to lack of pupil punctuation.

19) Relationship of the type-token ratio in oral and written language.—There is close similarity among the original story, oral samples and written samples. However, this measure is invalid because of the high rate of repetition of some words in the English language. A pupil who writes one sentence will obtain a higher type-token ratio than an author who writes a book. It could be a valid measure if an identical number of words were allowed for each sample.

20) Relationship of the Lorge grade rating in oral and written language.—The low degree of relationship as indicated in Table II is to be expected because of the dependence of the Lorge rating on sentencing.

SUMMARY OF ANALYSIS OF LANGUAGE SAMPLES

It would seem, from material presented in the preceding section, that there is substantial degree of relationship among reading material, oral language samples and written language samples of fifth- and sixth-graders in the following: verbal memories evoked, number of words used, number of different words employed, number of hard words, number of phrases and degree of subordination. Further, in situations similar to this study, such factors may be reliably measured. If such measures are used, different expectations or norms must be established for all factors except degree of subordination.

Factors considered in the preceding section which indicate some degree of relationship among the three media are: number of minutes for reading and reproduction, number of sentences, and type-token ratio. These have not proved suitable as measures of all three types of language behavior. While the number of

TABLE IV.—MEANS, STANDARD DEVIATIONS AND CRITICAL RATIOS OF DIFFERENCES BETWEEN THE MEANS OF ORAL LANGUAGE SAMPLES AND WRITTEN LANGUAGE SAMPLES FOR SEVENTY PUPILS IN EACH OF GROUPS 'A' AND 'B'

	Group A		Group B		Critical Ratio
	Mean	SD	Mean	SD	
Oral reproductions					
Total number of words	272.84	120.91	309.58	124.64	1.77
Number of different words	111.38	38.17	119.40	39.73	1.22
Number of hard words	17.41	8.28	19.10	7.57	1.26
Number of phrases	18.78	10.02	20.76	10.01	1.16
Number of sentences	13.64	6.38	16.78	7.06	2.76
Number of run-on sentences	2.00	1.80	1.91	1.83	.28
Number of incomplete sentences	.33	.67	.26	.55	.26
Degree of subordination	.31	.10	.32	.09	.91
Number of verbal memories	34.18	15.59	35.84	15.26	.64
Average sentence length in words	20.79	5.85	19.20	5.74	.63
Type-token ratio	.44	.09	.42	.06	1.95
Lorge rating	4.58	.35	4.42	.38	.14
Written reproductions					
Total number of words	188.63	72.72	204.33	88.49	1.15
Number of different words	91.30	27.48	97.01	33.98	1.09
Number of hard words	12.97	5.46	13.80	7.01	.78
Number of phrases	15.17	7.32	15.77	8.29	.45
Number of sentences	10.73	6.36	11.97	5.58	1.22
Number of run-on sentences	1.14	1.21	1.51	1.50	1.61
Number of incomplete sentences	.30	.66	.58	.94	1.88
Degree of subordination	.33	.11	.31	.09	1.23

TABLE IV.—(*Continued*)

	Group A		Group B		Critical Ratio
	Mean	SD	Mean	SD	
Number of verbal memories	30.01	12.66	30.76	14.45	.32
Average sentence length in words	26.00	30.31	19.71	10.82	1.64
Type-token ratio	.51	.07	.50	.08	.34
Lorge rating	5.04	1.84	4.58	.69	1.98

minutes for oral and written reproduction are positively related, there is little relationship of reading time with oral and written reproduction time. Differences of opinion as to what constitutes a sentence invalidates this as a measure. The type-token ratio is not a valid measure unless all samples are equated for number of words.

Factors which indicate little relationship among the media or are not applicable to more than two of the media are: average sentence length, the Lorge rating, mispronunciations, repetitions and excessive phrasing. Average sentence length and Lorge rating are suitable measures, but both are dependent on the definition of a sentence.

COMPARISON OF THE TWO EQUATED GROUPS

Statistical evidence of the result of comparison of oral and written language performance on the basis of order or presentation is given in Table IV. Group A first reproduced the story orally, then in writing. This order was reversed with Group B.

The only significant difference is in the number of sentences in oral samples. The writer is of the opinion that pupils who have just read the story rush through oral reproduction in the hope that facts will not be forgotten. Those who write the story prior to oral reproduction have undergone a sufficient time-lapse to assure that facts still remembered will remain so for a period of time.

Although no significant differences exist, with the exception of oral sentencing, the general trend as shown in Table IV is interesting. On those measures which were accepted as valid in the preceding section, Group B is superior in all but degree of subordination. Thus, according to evidence from the present study, pupils who write the story before telling it use, in both oral and written samples, a greater number of words, a greater number of different words, a greater number of hard words, a greater number of phrases and a greater number of verbal memories.

CONCLUSIONS

1) The level of development in the three language arts appears to be in the order: material read, oral language and written language. However, this result in a reproductive situation does not apply with equal validity to any other combination of language activities because the assumption is that the material read is the standard.

2) Varying the order of oral and written reproductions does not affect the quality of the samples, except for the number of oral sentences, to a significant degree. However, there is a general trend toward superior language usage in both oral and written samples when written reproduction is performed first.

3) The level of development of oral and written language is more dependent upon the level of the material read than upon the reading level of the pupils according to evidence of this study. The average level of maturity for oral language samples is identical with the difficulty of the passage read, both with Lorge rating of 4.5. Further experiments using reading material on various levels of difficulty would be necessary before this statement can be made with any degree of certainty.

4) Level of development revealed by one measure is not comparable with that revealed by another. Studies comparable to the present one have not been sufficient to establish levels of comparison among the measures. For example, it is not possible to state that one sample containing fifty more words but five fewer phrases than another sample is of less, equal or greater degree of language maturity.

5) Measures which appear suitable as multiple-measures. . . . which could be applied to material read, spoken and written in situations and with material comparable to the present study

are: verbal memories, number of words, number of different words, number of hard words, number of phrases and degree of subordination. The Lorge formula may prove to be an acceptable multiple-measure although such was not clearly the case in the present study. All of these measures except degree of subordination would require separate norms for the three media.

6) Assuming that the measures are suitable multiple-measures, the level of maturity of development in each factor could be determined only by further investigations. Comparable studies are not sufficient to establish levels of development. However, the present study indicates that, in similar situations, the mean number of written-language words may be approximately two-thirds that of oral reproduction; number of different words in oral samples may be about one-third the number of words in reading material; and the number of hard words one-tenth the total number of words, with written samples comparable on a reduced scale. There is further indication that in similar situations pupils may be expected to produce approximately one-third of the facts which they have read orally and that the ratio of subordinate clauses to total number of clauses used may be approximately one-to-three.

7) The hypothesis of Hahn(*9*), that excessive phrasing in oral reproduction is a habit caused by nervousness or excitement and tending to persist, appears to be substantiated here to the extent that this factor is related to reading and oral reproduction. The same relationship seems to be true for repetitions in oral reading and oral reproduction.

8) From evidence of the present study, the best single index of measurement in reading material, oral and written language samples would appear to be the degree of subordination.

BIBLIOGRAPHY

1) M. V. Bear. "Children's growth in the use of written language," *Elementary English Review*, 16:312–319, December, 1939.

2) P. P. Bushnell. *An Analytical Contrast of Oral with Written English.* (Teachers College Contribution to Education, No. 415.) New York: Bureau of Publications, Teachers College, 1930, pp. 87.

3) E. Dale. "A comparison of two word lists," *Educational Research Bulletin*, 10:484–489, December, 1931.

4) E. A. Davis. "The subordinate clause in oral and written language," *Child Development*, 12:333–338, 1941.

5) C. W. Dow and S. R. Papp. "The Relation of Reading Ability and Language Ability to Speech Ability," *Speech Monographs*, Published by the National Association of Teachers of Speech, 10:107–108, 1943.

6) A. M. Edwards. *Alphabetical Index of Occupations and Industries.* Washington: United States Government Printing Office, 1940, pp. 607.

7) A. I. Gates. "Character and Purposes of the Yearbook," *Forty-eighth Yearbook of the National Society for Study of Education*, Part II, 1949, pp. 1–9.

8) W. S. Gray. *Studies of Elementary-School Reading Through Standardized Tests.* Supplementary Educational Monographs, No. 1, University of Chicago Press, Chicago, 1917, pp. 157.

9) E. Hahn. "The speech of first-grade children in audience situations," *Elementary English*, 25:39–46, January, 1948.

10) F. Heider and G. M. Heider. "Comparison of sentence structure of deaf and hearing children," *Volta Review*, 43:364–367 and 406, June, 1941, 536–540 and 564, September, 1941, 599–604 and 628–630, October, 1941.

11) G. Hildreth. "Reading Progress in Grades II and III," *Forty-eighth Yearbook of the National Society for the Study of Education*, Part II, 1949, pp. 54–126.

12) W. C. Hoppes. "Some aspects of growth in English expression," *Elementary English*, 10:67–70 and 121–133, March and May, 1933.

13) E. Horn and J. F. Curtis. "Improvement of Oral Reading," *Forty-eighth Yearbook of the National Society for the Study of Education*, Part II, 1949, pp. 254–265.

14) W. Johnson and Others. *Studies in Language Behavior*, Psychological Monographs, 56:No. 2, 1944, pp. 111.

15) L. L. La Brant. *Studies of Certain Language Developments of Children in Grades Four to Twelve*, Genetic Psychology Monographs, 14:387–494, November, 1933.

16) B. K. Lemon and G. T. Buswell. "Oral and written expression in grade IX," *School Review*, 51:544–549, November, 1943.

17) I. Lorge. "Predicting reading difficulty of selections for children," *Elementary English Review*, 16:229–233, October, 1939.

18) I. Lorge and L. Kruglov. "The relationship between the readability of pupils' compositions and their measured intelligence," *Journal of Educational Research*, 43:467–474, February, 1950.

19) E. G. Mathews, R. P. Larsen, and B. Gibbon. "Experimental investigation of the relation between reading training and achievement in college composition classes," *Journal of Educational Research*, 38:499–505, March, 1945.

20) P. McKee. *Language in the Elementary School*. Boston: Houghton Mifflin Co., 1939, pp. 591.

21) L. J. Rossignol. *The Relationships Among Hearing Acuity, Speech Production, and Reading Performance in Grades 1A, 1B, and 2A.* Teachers College Contribution to Education, No. 936. New York: Bureau of Publications, Teachers College, 1948, pp. 50.

22) D. H. Russell. "Evaluation of Pupil Growth in and Through Reading," *Forty-eighth Yearbook of the National Society for the Study of Education*, Part II, 1949, pp. 284–301.

23) D. H. Russell, C. M. McCullough, and D. Gates. *Trails to Treasure.* Boston: Ginn and Co., 1949, pp. 45–56.

24) F. J. Schonell. "Diagnostic Tests in English," *The Yearbook of Education.* London: Evans Brothers, 1935, pp. 963.

25) ———. *Backwardness in the Basic Subjects.* London: Oliver and Boyd, 1942, pp. 560.

26) J. C. Seegers. "Form of discourse and sentence structure," *Elementary English*, 10:51–54, March, 1933.

27) M. J. Stormzand and M. V. O'Shea. *How Much English Grammar?* Baltimore: Warwick and York, 1924, pp. 224.

28) E. L. Thorndike. "Reading as reasoning," *The Journal of Educational Psychology*, 8:323–332, June, 1917.

29) A. F. Watts. *The Language and Mental Development of Children.* London: George G. Harrap and Co., Lt., 1944, pp. 354.

30) Z. E. Wiswall. "A study of sentence structure in eighth-grade composition," *Elementary School Journal*, 26:441–448, February, 1926.

IMPROVEMENTS IN READING RATE AND COMPREHENSION OF SUBJECTS TRAINING WITH THE TACHISTOSCOPE

HENRY P. SMITH and THEODORE R. TATE

The University of Kansas

Considerable interest has been aroused within the past five years by reports of research and statements of theory concerning gains in both reading speed and comprehension, and in some cases in other perceptual processes, following tachistoscopic training with digits or other materials.

Projectors equipped with tachistoscopes and instruments designed to present printed material at controlled rates of speed have been used alone or in combination with each other in numerous programs attempting to improve the reading performance of adult subjects. A number of pieces of equipment of both types are being offered for sale with the implication, at least, that a means for quick improvement in reading has been discovered. Data concerning the results obtainable have been moderate in amount.

The present article is a report of one attempt to gain information concerning the amount of improvement in adult reading ability which might accompany the use of such equipment. Data are presented concerning possible differences in personality and intelligence test scores between subjects who made large and small gains in ability to read material presented by means of the rate-controller; and, test scores are reported for those subjects who persisted in the training program well beyond their original objective of thirty-five sessions and those who dropped from the experiment before reaching their goal. In addition, the relationship between training and changes in test score on the Minnesota Clerical Test will be examined. This latter test was used in an attempt to examine the hypothesis suggested by Renshaw and others that tachistoscopic training may lead to a general improvement in form perception and that this improvement may transfer to reading and other special cases of form perception.

METHOD AND MATERIALS

Two reading-rate controllers manufactured by the Three Dimensional Company of Chicago and two standard 2″ × 2″ slide projectors equipped with Wollensak No. 3 Alphax shutter permitting exposures from one to 1/100 second were available for the experiment. Each projector was equipped with a Selectron (automatic slide changer) and a number of trays for use with the Selectron. This equipment made possible an efficient plan for progressive work, eliminated the mixing of slides, and allowed the subject to give continuous attention to the training material. Each tray contained thirty (five to nine digit) number slides, and different trays were used for each day's work. From thirty to sixty slides were used for each training session. After the first three sessions a shutter speed of 1/100 second was used throughout the experiment. The number of digits on the slides used by each subject was increased as soon as he became able to achieve accuracy on about eighty per cent of the slides he was using.

Eighteen college students volunteered to take extensive preliminary and end-tests and to participate in a minimum of thirty-five training periods of fifty minutes each. One-half of each training period was spent flashing digits onto a small screen with the tachistoscopic projector, reporting the digit observed, and checking the correctness of the response. The other half of the period was devoted to the reading of material on the controller. The speed of the controller was adjusted by the subject who was instructed to increase the speed as he felt he could do so and still continue to read the selection with understanding. Popular novels chosen by the subject were used as practice material. The difficulty of the novels varied from fifth to eighth grade. The subjects met daily, five days per week, with one of the authors present in the laboratory.

In addition to the record of the speed at which practice material was read on the rate-controller, measures of reading improvement consisted of weekly tests selected from *Penguin Island*, by Anatole France (tenth- to twelfth-grade difficulty, Dale-Chall formula). The Smith-Moler Test of Reading Effectiveness (Advanced Form) B, C, or D was given as a pre-test, on the thirty-fifty session and/or at the end of training.[1] All tests were administered as in regular test situations and were not presented by means of the controller.

The procedure for administering the weekly test was to allow the subject to read one or two chapters as a warm-up and then to read, for test purposes, a two to four thousand word continuation of the material under timed conditions. A short test was given over the selection after it was read. The Smith-Moler test was used to measure reading rate and comprehension on selections of three levels of difficulty.

Other data available concerning the subjects included pre-training scores on the Wechsler-Bellevue Intelligence Scale, Form II, and the Minnesota Multiphasic Personality Inventory. In addition, the Minnesota Clerical Test was given before training, after the thirty-fifth session, and/or at the end of training.

RESULTS

Eighteen subjects completed twenty-five or more training sessions, thirteen completed thirty-five or more sessions, six completed fifty or more, and two completed seventy sessions.

Table I gives the average speed at which the subjects were operating the reading rate controller on the first and fourth meetings and on every eighth meeting thereafter. While a record of the controller speed was made at each meeting, the points reported are sufficient to indicate the increased speed observed in use of the reading rate controller.

It will be noted that speed of reading on the reading rate controller rose steadily and does not appear to have reached a maximum by the sixty-eighth session.

The results on the weekly tests are shown in Table II. It appears that in the case of those subjects who continued training for thirty-five periods or more, substantial gains were made with-

---

[1] The Smith-Moler Test was developed for another study in reading improvement and is not yet in published form. Each of its four forms contains three reading selections of approximately 2400 words, 1400 words, and 1200 words of sixth-grade, college freshmen, and graduate level difficulty, respectively. The subject is allowed a specified time for reading each selection. When time is called he records the amount read. Per cent of comprehension is obtained by dividing the number of correct answers by the number of questions which could have been answered with the information obtainable in the portion of the selection which was read.

TABLE I.—MEAN READING RATE CONTROLLER SETTING FOR THE FIRST AND FOURTH SESSION AND EVERY EIGHTH SESSION THEREAFTER

Number of Training Periods	Number of Subjects	Training Sessions									
		1	4	12	20	28	36	44	52	60	68
25 or more	18	474	587	708	927						
35 or more	13	459	550	707	894	1127					
50 or more	6	457	556	749	921	1169	1481	1246			
70	2	424	380	453	507	1081	1543	1290	1556	1675	2120

out significant comprehension losses. Two subjects who continued training until the end of the twelfth week (sixty periods) had on that test an average speed score of 917 words per minute with seventy per cent comprehension as compared to their mark at the end of one week of 291 words per minute and 67.5% comprehension.

TABLE II.—READING SPEED AND COMPREHENSION SCORES AT THE END OF THE FIRST AND SECOND WEEK AND EVERY TWO WEEKS THEREAFTER

Number of Training Periods	Number of Subjects	Measure Used	Weeks						
			1	2	4	6	8	10	12
25 or more	18	speed	364	335	433				
		comp	80	70	72.5				
35 or more	13	speed	343	317	405	580			
		comp	80.7	70	75.7	71.9			
50 or more	6	speed	366	364	497	615	688	630	
		comp	81	68.3	77.5	75	74	69	
70	2	speed	291	307	476	728	619	854	917
		comp	67.5	60	75	60	55	55	70

A summary of the scores made on the Smith-Moler Test by the thirteen subjects who continued training for thirty-five or more periods is presented in Table III. It will be seen that an improvement of approximately fifty per cent in rate of reading on all three sections of the test was made by the thirty-fifth session but that in all cases this was accompanied by some drop in comprehension score.

TABLE III.—SCORES ON THE SMITH-MOLER TEST OF READING EFFECTIVENESS BEFORE AND AFTER THIRTY-FIVE SESSIONS OF TRAINING WITH THE TACHISTOSCOPE AND READING RATE CONTROLLER
(N = 13)

Level of Difficulty	Pre-Test		35th Session	
	Rate	Comprehension	Rate	Comprehension
Seventh Grade	302	91	447	76.6
College Freshman	179	83	312	73.9
College Graduate	178	75	239	69.2

Six subjects continued for fifty or more periods of training and made a somewhat greater improvement in speed (one hundred per cent on the College Freshman level material). They too suffered some drop in comprehension score except that a slight although probably not statistically significant gain was made in comprehension from the thirty-fifth to the fiftieth session on two sections of the test.

Eye movements of the subjects were photographed by means of the Ophthalmograph at the beginning of training, at the thirty-fifth training session, and/or at the end of training. The results of an analysis of the photographs are presented in Table IV.

The number of fixations necessary for each hundred words appears to show a regular and, in most cases, a substantial drop and the number of regressions is cut nearly in half.

An examination was made of the scores on the Minnesota Multiphasic Personality Inventory in an effort to determine differences in test patterns between subjects who continued training

TABLE IV.—OPTHALMOGRAPHIC RECORD OF SUBJECTS BEFORE AND AFTER TRAINING[1]

Sessions	N	Rate	Comp[2]	Span[3]	Fix[4]	Regress[5]
25	18	Pre 425	73.3	3.02	34.	4.4
(or more)		Post 518	81	3.96	27.8	2.5
35	13	Pre 400	69.6	3.11	33.2	4.46
(or more)		Post 540	88.4	4.12	26.8	2.15
50	6	Pre 370	65	3.16	32.5	3.3
(or more)		35th 559	83.3	4.	26.8	1.8
		Post 545	75	3.81	29.1	2.1
70	2	Pre 260	65	2.89	35.	6.5
		35th 633	87.5	4.06	24.	2.5
		70th 600	80	4.35	23.5	2.5

[1] Reading standard paragraphs supplied by American Optical Company.
[2] Comprehension in per cent.
[3] Span of recognition in words.
[4] Fixations per one hundred words.
[5] Regressions per one hundred words.

for fifty periods or more and subjects who stopped short of their previously agreed upon goal of thirty-five sessions. The number of cases involved does not justify more than a few tentative observations. There were four males who stopped training short of the goal and two males and two females who continued training well past the goal. The four members of the group stopping short of the goal tended to exhibit defensive reactions (higher K score—66 vs. 59.5) and their scores indicate that they appear to feel that they are discriminated against[2] (higher Pa score—61 vs. 51.5). The male subjects remaining beyond their original goal showed greater anxiety (higher D score—76 vs. 57) although the average D score for the two female subjects was 50 and the males in this group showed more deviate patterns in a neurotic direction and more femininity of interest (MF score 77 vs. 66) than did the members of the group who failed to reach the goal. About the only characteristic of the female profiles which distinguished them from normal was a tendency to over-anticipate.

---

[2] The writers are indebted to Dr. William Cottle, assistant director of the University of Kansas Guidance Bureau, for such interpretations of scores from the Minnesota Multiphasic Personality Inventory and Minnesota Clerical Test as are found in this article.

Comparisons were made also between the Multiphasic scores of the four subjects who made the greatest improvement in reading rate controller speed from the first to the thirty-fifth session (all males) and the three males and one female who made the least improvement. The group making small gain showed greater anxiety, (higher D score—62 vs. 55) psychasthenic behavior, (Pt score—64 vs 52) and tendencies toward paranoia (Pa 56 vs. 51). These tendencies may have operated as inhibiting factors to prevent increments in reading rates. This group appeared to have more emotional distractions than did the group making the largest gains. The average MF score for the three males making largest gains was 78 and for the four making smallest gains it was 71. It may be worth noting that ten male subjects who continued training for thirty-five periods or more made an average MF interest score of 71.6.

The Minnesota Clerical Test scores of the four subjects who attended fifty to seventy training sessions showed a rise from 55.1 on the numbers and 68.7 on the names portion to 76.5 on the numbers and 80.7 on the names. The four subjects who trained for approximately twenty-five sessions made an initial score of 63.5 on the numbers portion of the test and 79 on the names portion and a final score of 77 on the numbers and 63.3 on the names. This loss by the 'under' group on the names portion may indicate an indifferent attitude in taking the final test. Otherwise it is difficult to reconcile the general steady improvement of the 'over' group and the drop by the 'under' group.

The Wechsler-Bellevue scores for the group making the largest gain in reading rate controller setting was 121.1 on the verbal, 126 on performance, and 126.2 on the full test. The average score of the four subjects making the smallest gain was 127 on the verbal, 117 on the performance, and 124.7 on the full scale. Thus, the verbal score of the group making the smallest gain was already markedly higher than was their performance score, while the reverse was true in the case of those making the largest gain. This would appear to indicate that the latter group had more potential for improvement in verbal skills. Had the Wechsler-Bellevue scores been obtained again at the end of the study it would be interesting to see if the latter group did now show a substantial gain in verbal score with a resultant increase in the general intelligence quotient.

The primary purpose of the Minnesota Clerical Test was to investigate the types of perceptual improvement other than reading which might accompany training with the tachistoscope. The results of this test as administered at various points in the training program are reported in Table V.

TABLE V.—PERCENTILE SCORES ON THE MINNESOTA CLERICAL TEST BEFORE TRAINING, AT THE END OF THIRTY-FIVE SESSIONS, AND/OR AT THE END OF TRAINING

Number of Training Periods	Number of Subjects		Pre-Test	35th Session	End-Test
25 or more	18	Numbers	60.7		73.9
		Names	68.2		75.2
35 or more	13	Numbers	58.5		71.3
		Names	69.0		75.0
50 or more	6	Numbers	57.16	69.91	77.58
		Names	72.41	80.58	82.83
70	2	Numbers	50.25	65.5	75.75
		Names	57.25	67.5	71.

Table V appears to indicate that increments in Minnesota Clerical Test scores did occur with continued training up to the end of the present experiment. Increases for the numbers portion of the test were considerably greater than for the names portion. The reason for this may lie in the rather direct use of numbers in the tachistoscopic portion of the training procedure and some failure of the gain in perceptual skill to transfer completely to a different type of task.

GENERAL COMMENTS

As shown by their settings of the reading rate controller and by their frequently expressed opinions, the subjects appeared to believe they were obtaining tremendous improvements in reading speed as a result of either the tachistoscopic training or the practice on the reading rate controller or from a combination of the two.

While the various tests employed indicated substantial improvement in reading rate, the improvement as measured by reading tests was not nearly as great as was shown on the controllers. While this may not be taken as conclusive evidence that reading speed and comprehension were not built to a high level as a result of the training nor that such improvement cannot be transferred to reading independently of the machine, it is clear that but a small portion of the indicated speed did transfer in the present experiment and this at some loss in the area of comprehension.

However, the indications from the comparatively low but actually rather high improvements obtained in test performance and in eye-movement records indicate that further experimentation may be of value. On the other hand, the results of this and other experiments indicate that as yet too little is known concerning the effect of such equipment on the reading ability of persons of various ages, degrees of intelligence, and varying types of personality patterns to warrant general use of the equipment in remedial reading programs.

MILTON GURVITZ. *The Dynamics of Psychological Testing.* New York: Grune and Stratton, 1951.

Although written primarily for clinical students, this book was inspired by a non-clinician's comment, "I'd really like to see some time how you clinicians use these tests." Gurvitz has provided a thorough demonstration of one style of clinical interpretation. He presents seventeen cases, selected from the patients tested during one month in a mental hospital. Typically, the case report includes a scored Rorschach protocol, Wechsler protocol, figure drawing with inquiry, a discussion of the dynamics of the case, the original psychological report with interpolated agreements or reservations by the therapist who later handled the case, a formal case history, and a final summary.

It is a mistake for Dr. Joseph Miller, who writes the foreword, to refer to this as research. It is essentially a report of clinical experience and opinion. There is no guarantee that Gurvitz did not select cases where his diagnostic procedures worked well, and so this provides no solid evidence of the validity of the specific procedures. The reports do show that tests can contribute greatly to the understanding of patients.

Gurvitz is an able interpreter of behavior. He studies his patient as a person, making as much note of his over-all behavior as of the test responses. He has a wise viewpoint regarding his tests, making excellent critical comments on some of the proposals of Rapaport, Wechsler, and the Rorschach sign systems. The book as a whole shows high level psychological skill in action. Gurvitz makes especially good use of the Wechsler and figure-drawing data.

Too often, Gurvitz writes statements which, read by themselves, would encourage mechanical and unintelligent use of tests. Once in a while he seems to say that a diagnosis is decided by some single response of the patient. He seems, despite his disclaimers, to use signs himself in diagnosis. For instance, on the subject of W:M in Rorschach, he says (p. 21): "If M predominates, then we have a surfeit of ability and creativeness but insufficient drive to project it out into the world." Gurvitz relies heavily on a Freudian terminology and way of thinking which does not seem to be an integral part of his diagnostic skill. His dogmatic and atomistic statements about particular indicators will blind some readers to the fact that his actual diagnosis is based on a thorough integration which allows each fact about the patient to add to the significance of each other.

The Rorschach interpretation suffers greatly from an incautious attitude. Gurvitz seems to accept almost every idea of Klopfer without reëxamination. The literature contains enough validation research by this time to suggest that many items in this interpretive system are questionable. The statement quoted about W : M ignores questions of unreliability. It does not admit that the relation between personality and action is so complex that not even the most valid statement will be true of all cases. It does not acknowledge that the current literature contains almost as many conflicting theories of the meaning of M as there are writers. It does not warn the reader that a test indicator which goes with some defect in a hospital population will also appear among any normal group tested, with no corresponding defect. Gurvitz probably makes a more flexible use of the Rorschach, and a more cautious use, than his generalizations will suggest to the reader.

It is perhaps too much to hope that writers like Gurvitz will follow each suggested interpretation with a statement of the percentage of cases for which the interpretation is valid, out of all those where the indicator appears. Until they do make statements in those terms, they can expect that non-clinicians will continue to regard clinical psychology as dogmatic and unsound.

This book can be used profitably with advanced students in clinical diagnosis by two types of instructor. The ones who wish students to be skeptical of over-enthusiastic and fine-drawn interpretations will find here examples worthy of critical attention. These reports are neither illogical nor lacking in insight. The reasoning is careful and the psychology insightful; criticism must therefore focus on the premises underlying the interpretation. The instructors who want to teach students to squeeze test protocols for all they are worth, without too much fretting about lack of validation, should also use this book. If students are to learn this brand of clinical psychology, they should be taught from a model as skilled as Gurvitz.

While Gurvitz' writing does not do justice to his acumen, his thinking about tests is in many ways sounder than that of some other clinicians and of those who look to a test for a score and nothing but a score. His concept is worth quoting:

" . . . Our present psychologic technics give us the maximum amount of information in a minimum of time and with a lessened degree of subjectivity. The tests which are used clinically . . . are considered to offer a wide range of possible environmental stimuli. Conflicts are aroused in miniature to see how they are handled or mishandled, anxiety is provoked, basic relationships to figures in the environment are encountered, phantasy is forced upon the patient—all to duplicate the world in miniature. The process of interpreting psychologic procedures is based upon the assumption that the person handles the microcosm of the testing situation in the same manner as he handles life." (p. 7)

LEE J. CRONBACH

University of Illinois

GEORGE G. THOMPSON. *Child Psychology: Growth Trends in Psychological Adjustment.* Boston: Houghton Mifflin Co., 1952, pp. 667. $5.50.

Child psychology has been an important special interest in the total domain of psychology since the earliest years of this century. In the post-war years the impression is easily gotten that this interest has decreased relatively, if not absolutely. Attention is being directed so much to industrial, social, and clinical problems of adults that the significance of the child is perhaps too much neglected. Yet, that there has been sound and important research with children, and that the developmental processes are of significance in understanding adult behavior, cannot be denied. The evidence for this statement is to be found in Thompson's book.

As an interpretive exposition of the status of scientific child psychology at mid-century, this volume will find an important place in the psychological literature. Systematically, the author moves from an introduction of child psychology as a scientific discipline, through a review of the behavior patterns of the newborn, to more extensive considerations of the processes of psychological growth and adjustment, the interactions of motivation and learning, personal and social adjustment, and finally a survey of the theories of personality organization. The organization is determined by psychological processes rather than chronological age groupings. Changes with age are pointed out, but the process of growth, rather than its stages, is constantly foremost in the argument. Research results and theoretical positions are expertly related to each other and woven into a coherent pattern.

The writing style is clear and readable. As a textbook for advanced undergraduates this volume is excellent; there might be some doubts of the author's claim that the contents "should be understandable to the average college freshman" even with "occasional assistance from the instructor." Outside of the college classroom there are many professional workers with children, and laymen—parents or others—who will find the reading of this book of immediate interest, and who can turn to it as a source of information for some time to come. The chapter bibliographies, totalling nearly one thousand references, alone are an important guide to the literature of the important and fascinating field of child psychology. C. M. LOUTTIT

University of Illinois

WILLIAM C. MORSE, FRANCIS A. BALLANTINE, AND W. ROBERT DIXON. *Studies in the Psychology of Reading*. Ann Arbor: University of Michigan Press, 1951, pp. 188.

Three research studies on eye movements in reading are reported in this volume. Morse's study deals with changes in eye movements when fifth- and seventh-grade pupils read materials which are easy, at grade and difficult. The descriptive science materials used were equated for difficulty by readability formulas. The eye-movement records revealed that seventh-graders were more efficient readers than fifth-graders, and that changes in difficulty of materials produced no significant changes in reading. Thus, eye-movement patterns tend to remain stable in various situations.

Ballantine measured oculomotor changes in silent reading at various grade levels from the second through the twelfth. The records revealed rapid improvement through Grade IV, little change from IV to VI, further gains through VIII and little growth thereafter. Again the eye-movement patterns did not change with change in difficulty of textural material.

Dixon's subjects, specialists in physics, history and education, read material in their own field and in the other fields. The subjets read material in their own field most efficiently. It is considered that familiarity of material is a factor in the reading performance but that different types of material do not produce different eye-movement patterns.

In all these studies the experimental designs are excellent, and the results represent worthy contributions. The reviewer is disturbed by the fact that in all three studies a change in difficulty or content of reading material brought no changes in oculomotor patterns. It is generally accepted that the most effective reader is the one who modifies his pace to fit the requirements of the content and the difficulty of the textual materials. Lack of such flexibility in readers seems unfortunate. Further research on this topic is desirable.

MILES A. TINKER

University of Minnesota

DOROTHY C. ADKINS AND SAMUEL B. LYERLY. *Factor Analysis of Reasoning Tests.* Chapel Hill: University of North Carolina Press, 1952, pp. iv + 122.

The authors factor-analyzed first a battery of thirty-eight selected Air Force tests, in order to obtain leads for selecting tests to go into their main experimental battery. The latter contained sixty-six tests, which were administered to a fairly representative sample of two hundred soldiers. Product-moment intercorrelations were computed, and sixteen factors were extracted by the complete centroid method and rotated as nearly as possible to an oblique simple structure.

The following reference factors were put in (two tests for each) and taken out again: verbal relations, perceptual speed, number, word fluency, space 1 (visualization of rigid figures under rotation and translation), speed of perceptual closure, ideational fluency, and space 2 (visualization of figures whose parts move in relation to one another).

Five reasoning factors were identified and described: perception of abstract similarities, hypothesis verification, flexibility of perceptual closure, deduction, and concept formation. Perception of abstract similarities is defined by high loadings on verbal classification tests and on analogies tests, both verbal and nonverbal. Hypothesis verification is defined essentially by high loadings on the progressive matrices test. Flexibility of perceptual closure is defined by high loadings on figure classification tests of the types which require allocation of a fairly large number of test items to one of two or three groups after the groups have been distinguished, and by high loadings on Gottschaldt Figure tests and others more or less similar. Deduction is defined mainly by high loadings on syllogisms tests and verbal analogies tests. Concept formation is defined most clearly by tests which require the examinee to supply a general name for a group of words or objects.

In the original analysis, sixteen factors were extracted, even though the common-factor variance was substantially exhausted after the extraction of the fourteenth. After rotation there were three factors which were uninterpretable and did not conform well to the criterion of simple structure. All of these factors contained reasoning tests of one sort or another with loadings above .30; five tests had their highest loadings on one or other of these 'residual' factors.

This study shows a remarkable number of points of disagreement with previous factor analyses. Since it is the first major study aimed specifically at determining the nature of the factors in the reasoning domain, some disagreement with the results of previous work is to be expected. However, the points of difference are so numerous and serious that to the present reviewer the substantive findings (the factors) must still be considered tentative, and their interpretations hypothetical.

It is possible that a re-analysis of these data might result in a large improvement in the interpretability of the results. The authors indicate that their subjects also took all tests of the Army Classification battery, and that the intercorrelations among the ten tests of that battery, as well as their cross-correlations with the sixty-six variables of the present study, were computed. The use of some at least of these additional test scores would undoubtedly improve the definition of some of the reasoning factors, and especially of some of the reference factors. It is also probable that in the case of these data a recomputation by principal axes or maximum likelihood would yield a sharp cut-off of common factor variance after the fourteenth or perhaps some earlier factor. A new rotation might then provide a much clearer picture. It is very much to be hoped that such a reanalysis will be made. The reviewer is impressed by the effort and ingenuity that have gone into this study, but disappointed by the inconclusiveness of the results. His own evaluation suggests that this inconclusiveness may possibly not be intrinsic to the data.

EDWARD E. CURETON

University of Tennessee

RUTH E. HARTLEY, LAWRENCE K. FRANK AND ROBERT M. GOLDENSON. *Understanding Children's Play.* New York: Columbia University Press, 1952, pp. 372. $3.50.

This book is the result of an exploratory study undertaken in 1947 and sponsored by the Caroline Zachary Institute under a grant from the National Institute of Mental Health. The purpose of the study was to determine the effect of play experience upon personality development of preschool and kindergarten children. This study of some one hundred eighty children from two to six years of age was organized and supervised by Frank; Hartley directed and conducted the project with the assistance of Mrs. Ellen Schindel. The New York State Mental Health Authority provided additional funds which enabled the directors of the exploratory study to discuss their findings and evaluate them with groups of teachers, nursery school directors and child center directors. Goldenson has taken Hartley's original manuscript which represented the most pertinent findings of the project and has condensed and revised it so that it could be of most use for directors, teachers, parents, and those concerned with the growing child and the promotion of his mental health.

The nine chapters in the book cover the following topics: dramatic play as a mirror of the child and as an instrument of his growth, importance of block play as an outlet for childhood expression, benefits of water-play, clay not only as a projective tool but as a raw material for construction, use of graphic materials as media for the child's expression of feelings, finger painting not only as a diagnostic device but as a means of creative expression, and lastly the combination of music and movement as a therapeutic device with children. Each chapter discusses pertinent previous studies, if any, gives a wealth of recorded observations of children using a particular play medium, includes interpretations of the anecdotes and concludes with helpful suggestions to teachers and those who come in contact with the preschool and kindergarten child. The appendix lists extensive suggestions for making observations and recording them; these are detailed enough to be used by the layman. Notes and bibliographic references are given by chapters separately. A fairly adequate index is included.

The reviewer's greatest criticism concerns the fact that the observations of the play activities of the children included in the study were not statistically summarized so that the reader could get some idea of the similarity and differences in behavior of the well-adjusted and the poorly-adjusted child to the same play media. Occasionally the similarity and difference of behavior between the inhibited and aggressive child in one media, say finger painting, is pointed out but the only proof offered for such similarity is another anecdote concerning another child. Throughout the book terms are used loosely with no attempt to define meaning. Unfortunately the book is marred by assertions made by the authors which occur on nearly every page, which are not verified, and which are then interpreted as being true for most children.

The diversity of authorship has led to discontinuity in thought. Repetitious phrases occur throughout the book and detract from what might have been an interesting and informative account of children's play experience in relation to personality growth.

The reviewer doubts that teachers, parents, and social workers will derive as much benefit from this book as is claimed. Teachers and social workers with psychological backgrounds may understand the implications of the findings, but parents will find it difficult to separate interpretation from fact, let alone make use of interpretation based on observations of the children used in this study.

R. Elizabeth Brown

University of Illinois

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 44 May, 1953 Number 5

VoL 44 No 4, 1953 Mcching

CONCEPT FORMATION OF WORK-STUDY SKILLS BY USE OF AUTOBIOGRAPHIES IN GRADE FOUR

WALLACE J. HOWELL

Principal, George M. Diven School
Elmira, New York

By definition, "a concept is an idea that includes all that is characteristically associated with, or suggested by, a term." A term, to become meaningful and applicable, must become a part of the child's sensory impressions and his thinking. If this is to be accomplished in the many areas of education, all terms must be presented so that they are understood, and suitably integrated into the learning experiences of the child. In the study here reported a program of work-study skills was decided on to provide the learning experiences, but the desirability of testing its effectiveness before it was put into operation was recognized. The improvement in performance could be ascertained by the use of an achievement test, but it was also desired to learn the extent to which the experiences would become internalized by the pupils. On the principle that individuals tend to report spontaneously on events involving a cognitive reorganization, a situation was provided for those who had the experiences to make such a report. Others who were familiar with all the activities involved but without the experiences with the work-study skills would be given a similar opportunity. The technique of measurement chosen was the pupil autobiography. The hypothesis, stated positively, was that the pupils who had had them, if the experiences were internalized, would reveal the fact through the number of items relating to the experiences appearing spontaneously in their autobiographies. An initial autobiography, an intervening period of training, and an added 'chapter' called for at the conclusion of the experiment should show significant differences. The items mentioned by fourth-grade children in their spontaneous autobiographies should also shed light on some of the problems hindering their adjustment and personality development and also reveal differences between boys or girls in this respect.

If the results were statistically significant, a trend would be shown which might aid in reducing the problems involved in teaching work-study skills. This becomes a factor of great importance, and the manner in which these skills are presented depends upon the techniques of informing, instructing, and teaching the child, as well as the manner in which the child gives back the information. This is based upon the assumption that the skill has reached the realm of conceptual reality as both the teacher's purpose and the child's purpose are consummated.

PROCEDURE

The research design for this study involved the use of the matched-group technique. An experimental group consisting of eighteen boys and twenty-five girls and, in a different school, a control group of the same number of boys and girls were formed by the matching process. This made a total of eighty-six children—thirty-six boys and fifty girls. These groups were matched on the basis of the following criteria: Age in months; intelligence quotients derived from the New California Short Form Test of Mental Maturity, Elementary '47, S-Form; the raw reading scores obtained from the Iowa Every-Pupil Test of Basic Skills, Test A, Form M; and sex. The method described by Peters and Von Voorhis *(2)* was used with these criteria in the matching process. A comparison of the mean and the standard deviation in each of the three criterion areas in the experimental and control groups, as given below, re-

Criteria	Experimental Group		Control Group	
	M	SD	M	SD
Age in months	116	7.05	115	9.78
Raw reading score	60	16.9	57	17.5
IQ	103	16.0	105	13.0

veals the accuracy of this matching process. The study was begun in September, 1949, and extended for nine months, ending in May, 1950.

In September, 1949, after the testing program and the equating process was completed, each child was asked to write his autobiography without using a guide or outline. Each autobiography was carefully read and all items mentioned by the children were tabulated. These items were categorized into ten major groups with the frequency of mention as follows: Favorite subjects (123); Liking for teacher (53); Parts of school building (30); Schools attended (25); Best liked books (17); Liking for schoolmates (17); Test items (14); Difficult subjects (9); Liking for flag (4); and Mother's criticism (1). As a result, the eighty-six children in the initial autobiographies mentioned a total of 293 items or an average of 3.5 items per child. The eighteen boys of the experimental group mentioned 71*[1] items, or twenty-four per cent of the total, while twenty-five girls of the same group mentioned 80* items or twenty-seven per cent of the total. In the control group the eighteen boys listed 50* items or seventeen per cent while the girls names 92* items or thirty-one per cent of the total. Percentagewise, very little difference appeared between the equated groups. The thirty-six boys mentioned 121* items or forty-one per cent of the total while the fifty girls mentioned 172* items or fifty-nine per cent of the total.

Besides the autobiographies the eighty-six children involved in the study were given the Iowa Every-Pupil Test of Basic Skills, Test B, Form M, in the fall and Form N again at the conclusion of the study to further measure the difference in achievement as a result of the year's work.

An intensive program of instruction was begun in the area of work-study skills in September, 1949. Twenty-three units of work were presented to the experimental group while their control counterparts were not exposed to this intense program. Each unit was carefully and thoroughly presented to the experimental group by the librarian. The teacher of the experimental group then utilized the content of the units and correlated it with all the subjects of the school curriculum. A careful record was kept of the number of times each unit was correlated with the curriculum. The twenty-three units included such topics as parts of a book; index; encyclopedia, both general and special; dictionary; atlas; *World Almanac*; features common to all reference books; outlining, including general technique as well as outlining from visual and auditory perception; and reading with emphasis upon various types applicable to different purposes of the reader, such as comprehension and skimming.

---

[1] The asterisks here indicate frequencies of mention for which reliabilities were found.

In May, 1950, after nine months of intensive work, the initial autobiographies were returned to the children with the simple request to add a chapter to the original. Again, the added chapters were carefully examined and the various items tabulated. Table I categorizes the children's responses according to the same pattern followed in the initial autobiography.

TABLE I.—NUMBER AND PER CENT OF ITEMS MENTIONED IN THE ADDED CHAPTERS OF THE AUTOBIOGRAPHIES BY EIGHTY-SIX CHILDREN IN GRADE IV IN THE EXPERIMENTAL AND CONTROL GROUPS BY SEXES. SPRING 1950

Categories of Items Mentioned	Experimental Group				Control Group				Total			
	18 Boys		25 Girls		18 Boys		25 Girls		36 Boys		50 Girls	
	N	%	N	%	N	%	N	%	N	%	N	%
Work study skills	74	35	114	54	6	3	16	8	80	38	130	62
Likes teacher	7	17	12	29	7	17	15	37	14	34	27	66
Favorite subjects	7	10	11	15	32	14	23	31	39	53	34	47
Countries and regions studied	8	42	6	32	1	5	4	21	9	47	10	53
Difficult subjects	3	43	4	57					3	43	4	57
Parts of building	1	20	2	40	1	20	1	20	2	40	3	60
Secondary items	5	21	4	17	6	25	9	37	11	46	13	54
Total	105*	28	153*	40	53*	14	68*	18	158*	42	221*	58

* Indicates frequencies of mention for which reliabilities were found.

The terms comprising the first category (work study skills) as given by the experimental group in Table I will be listed in descending order, with the frequency of mention given after each term, since our major hypothesis is the formation of concepts of work-study skills. These terms are: Dictionary (22); Encyclopedia (21); Trip to city library (19); Likes school library (12); How to study (11); *World Almanac* (11); Atlas (10); Maps (9); Index (5); Card catalog (5); Globe (4); Movie on how to study (4); and Learned more than ever before (3). The remaining fifty-two items mentioned could be considered adjustive in nature because such items as likes school and happy in grade, lots of fun, likes work of grade, exciting year, enjoyed extra hard work, likes classmates, year passed by fast, etc., indicate that the social climate of the classroom brought about by the year's work impressed the children considerably.

The effect of this intensive work during the nine months is revealed by the fact that, unsolicited, these children in the experimental situation listed twenty-seven different terms one hundred eighty-eight times in the category of work-study skills, or seventy-three per cent of all items mentioned in the added chapter by this group. This definitely demonstrated that these items became a part of the child's thinking since he was able to recall and use these terms. The children in this group average 6.0 items per pupil on the added chapter.

In contrast, the control group mentioned twenty-two items definitely classified in the work-study skills category. These were eleven per cent of all items mentioned in the added chapter. The average number of items mentioned was 2.8. Aside from the work-study skills items, all other items mentioned in the other categories follow the general pattern revealed in the initial autobiographies.

Considering the total pattern of Table I, the girls again mentioned a greater percentage of the items than did the boys—fifty-eight per cent and forty-two per cent, respectively. The added chapters revealed certain curriculum areas dwelt upon during the year which were new to the fourth-grade children. The experimental group mentioned more areas more often than the control group, and the children mentioned their favorite subjects more often than their most difficult subjects.

At the close of the experiment the Iowa Every-Pupil Test of Basic Skills, (Test B, Form N) was again administered for the purpose of checking the results obtained in Work-Study Skills with a standardized instrument. These results, compared with those of the autobiographies, should throw additional light upon the significance of the results obtained. Table II gives the mean results of this test in the fall of 1949 and the spring of 1950.

Many interesting trends are revealed by a study of the comparative results of the initial and final testing program given in this table. The total picture favored the experimental boys over those of the control by .9 of a year, whereas the experimental girls surpassed those of the control by .8 of a year. The greatest growth occurred in the area of Use of References, followed closely by Alphabetization, Use of Index, and Use of the Dictionary. In Map Reading the experimental group was surpassed by the control group.

Table II.—Mean Gains in Grade Equivalents Made in September, 1949, and May, 1950, on Test of Work-Study Skills by Experimental and Control Groups by Sexes

Test Areas and Date of Testing	Experimental Group		Control Group	
	Boys	Girls	Boys	Girls
Map Reading				
September 1949	3.9	3.3	3.5	3.3
May 1950	4.9	4.5	5.0	4.3
Gain	1.0	1.2	1.5	1.0
Use of References				
September 1949	3.1	3.3	3.7	3.5
May 1950	5.8	5.0	3.8	4.0
Gain	2.7	1.7	.1	.5
Use of Dictionary				
September 1949	3.5	4.0	3.5	3.5
May 1950	5.0	5.0	4.4	4.5
Gain	1.5	1.0	.9	1.0
Use of Index				
September 1949	3.2	3.5	3.2	3.5
May 1950	4.9	5.0	4.5	4.4
Gain	1.7	1.5	1.3	.9
Alphabetization				
September 1949	3.3	3.3	3.4	3.0
May 1950	6.2	5.7	4.2	4.4
Gain	2.9	2.4	.8	1.4
Total				
September 1949	3.5	3.4	3.5	3.4
May 1950	5.3	5.1	4.4	4.3
Gain	1.8	1.7	.9	.9

Considering 4.9 as the norm for the May, 1950, testing program, the experimental group surpassed this norm in all areas except that of Map Reading. In this area the girls fell below the norm.

The control group was below this norm in all areas except in Map Reading where the boys were one month ahead of the experimental. The above results of this standardized measure are highly significant, making further discussion unnecessary.

In order to ascertain whether or not the results of the autobiographies are significant as far as the work-study skills are concerned, Chi-square was used to test the null hypothesis, which assumes that there is no difference between the two populations and that there is no difference between the responses of the boys and girls in the selected sample. A contingency table was set up for both the boys and girls separately and in toto involving two columns consisting of the experimental and control results, and four rows including the initial and added chapters of the autobiography items. The formula shown by McNemar (*1*) was used to determine the value of X^2.

This formula was applied to the tabulated results found in Table I for the added chapter and a breakdown of the results given above for the initial chapter. The numbers above, identified with an asterisk, were used in this formula. It was found that the total results were highly significant, with X^2 yielding a p much less than .001, and the null hypothesis was rejected, indicating that a real difference exists in these experimental findings. As far as the girls were concerned X^2 yielded a p much less than .001, indicating a significant difference. Likewise, the null hypothesis was rejected. The results for the boys yielded a p of approximately .60 and the null hypothesis was not rejected. All in all, this indicated a real sex difference existed favoring the girls.

RESULTS

From the above data the following results can justifiably be listed:

1) The number of items mentioned by the experimental group in both the initial and added autobiographies, disregarding sex, totaled four hundred nine or sixty-one per cent as compared with the two hundred sixty-three items or thirty-nine per cent mentioned by the control group. This difference is significant at less than the one per cent level of confidence as revealed by Chi-square and can be attributed to the year's intensive work in the work-study skills area.

2) Of the total six hundred seventy-two items mentioned in both the initial and added phases of the experiment, the girls mentioned three hundred ninety-three items or fifty-eight per cent of the total as compared with the two hundred seventy-nine items or forty-two per cent for the boys. Chi-square showed the results for the girls to be highly significant with a p of less than .001 while the results for the boys yielded a p of .60.

3) A decided sex difference occurred in this experiment favoring the girls from the standpoint of items mentioned in the autobiographies.

4) In the initial autobiographies the experimental group averaged 3.51 items per pupil and the control group averaged 3.30 items per pupil. However, in the added chapters the forty-three children in the experimental group averaged 6.0 items per pupil compared to an average of 2.81 items per pupil in the control group. This indicates a definite trend favoring the experimental procedures.

5) The average number of items mentioned by the experimental group in both the initial and added chapter was 9.51 as compared with 6.1 for the control group. The average number of items mentioned by all the eighty-six children was 7.81.

6) The experimental group evidenced desirable concept formation of work-study skills in listing one hundred eighty-eight terms in their added chapter which was seventy-three per cent of the total items mentioned by that group. On the other hand, the control group listed twenty-two items or eighteen per cent of the total mentioned by them. This great difference seems clearly to be due to the intensified work in the area of work-study skills and the corresponding concept formations.

7) The fifty-two items classified as adjustive in nature definitely relate to the social climate of the classroom. These adjustive terms pertain to the personality development of the children and those possessing these concepts were, without a doubt, definitely impressed and better adjusted.

8) A comparison of differences in gains in grade equivalents between the experimental and control groups indicates that the greatest growth occurred in the area of use of references, followed by alphabetization, use of index, and use of dictionary. All of these areas received special emphasis and the gains further indicate that

the experimental procedures aided in the formation of desirable concepts in the skill areas as postulated.

9) In the re-testing both boys and girls of the experimental group equaled or exceeded the norm of 4.9 in all areas except that of map reading. The boys and girls of the control group were below the grade equivalent norm of 4.9 in all areas of the test except map reading. Again the major hypothesis is satisfied.

10) In the fall testing of work-study skills all children fell below the grade equivalent norm of 4.0 except the experimental girls in the use of dictionary area.

11) The multi-sensory experiences used here motivated the acquisition of the concepts of work-study skills through use of the concrete approach.

12) What has been demonstrated as occurring during the nine months of this study attests to the postulated major hypothesis concerning concept formation of work-study skills.

CONCLUSIONS

Autobiographies, as used here, definitely reveal satisfactory evidence that this concept formation is more evident among the girls than the boys. Expecially is this true when they are written without benefit of an outline or guide. Furthermore, autobiographies furnish a great amount of information and proper dissemination of this information can do much to indicate and eliminate maladjustments. The autobiography is a tool whose potential worth in the area of child development has been insufficiently explored, but the spontaneity of terms given by the children and the indication of adequate conceptual formation of skills being explored make this approach highly significant in measuring the effects of learning experiences upon individuals over an interval of time.

REFERENCES

1) Quinn McNemar. *Psychological Statistics*. New York: John Wiley & Sons, 1949, p. 207.

2) Peters and Von Voorhis. *Statistical Procedures and Their Mathematical Bases*. New York: McGraw-Hill Co. 1940, pp. 448–451.

THE DEVELOPMENT OF A TEST TO MEASURE THE INTENSITY OF VALUES

JOSEPH E. SHORR

Los Angeles, Calif.

Empirical evidence of the importance of value systems as an organizing and motivating factor in behavior has been well accepted. Following its development in the early thirties, Allport and Vernon's Study of Values Test (*22*) has been the main instrument used to measure six systems or patterns of values; namely, the Theoretical, Social, Political, Economic, Aesthetic, and Religious. Since that time almost fifty articles have been published showing the importance and stability of the value concept.

The Allport and Vernon Study of Values utilizes relative scales of a forced-choice type and this results automatically for some of the six scores to be high and some of the scores to be low. A higher score on one type of value makes for a lower score on some other type of value. However, it is conceivable that an individual may be actually low on all the scales or actually high on all the scales or possibly medium on all the scales or other various combinations. Moreover, more careful examination revealed that a beginning student in college physics may often secure approximately the same raw score on the Theoretical scale as a physicist who was keenly interested in his field. Furthermore, people who were not intensely interested in Aesthetics would sometimes approximate in score those strongly interested in Aesthetics. In a similar manner the Social, Political, and Religious scales tended not to differentiate those who valued quite strongly from those who had better than average interest and scored equally high because of the relative scales. It appears that the higher and lower extremes of a representative sample were not differentiated and that an 'artifical' interdependency of each score upon each other existed.

Allport, Vernon and Lindzey (*23*) have the following to say in the newly revised edition of the *Study of Values*: "In interpreting the results, therefore it is necessary to bear in mind that they reveal only the relative importance of each of the six values in a given personality, not the total amount of 'value energy' or drive possessed by an individual. It is quite possible for the highest value of a generally apathetic person to be less intense and effective than the lowest value of a person in whom all values are prominent and dynamic."

It was felt that this defect could be avoided by measuring the intensity of each scale of values from a practicable minimum of intensity to a practicable maximum of intensity.

PROCEDURE

In order to make a new scale that would encompass the entire range of value-interest intensity as nearly as can be approximated, the following steps were taken into consideration:

1) A broad matrix of items were gathered ranging from, for example, "Avoid social contacts," to "Work with labor and management to help solve their conflicts," in the Social Scale with sufficient items to cover about nine intermediary steps of intensity. In addition, one hundred and fourteen items that in the opinion of the author showed various degrees of intensity in Theoretical interests were gathered and collected on cards. One hundred and thirty-one Social-value questions of different intensities were so secured. One hundred and thirty Aesthetic items of various intensities, and one hundred ninety-two Economic-Political items varying in intensity were also typed onto cards.

2) Because of empirical usage it was decided that four of the six scales be retained and that, in doing so, the Economic and Political scales should be combined. It was felt that the Religious scale could be eliminated because as Super (*21*) says, "The religious values scores do not, in some cases represent more than the lip service of immature persons who have as yet experienced neither deep religious feeling nor intellectual doubts concerning religion."

3) The items for each scale were then rated on an eleven-point scale ranging from a negative avoidance level to a level of maximum intensity. The technique employed was essentially the Thurstone equal appearing method of scale construction.The items were rated by eleven raters, all of whom were familar with value theory and value tests. They were given no other instructions but to rate them in intensity from low to high on an eleven-point scale. Each scale was described to them by prepared statement as follows:

> *Theoretical.*—A high score indicates that the individual prefers and considers most worth while those activities which involve a problem-solving attitude and are related to investigation, research, and scientific curiosity.

Economic-Political.—A high score indicates that an individual prefers and considers most worth while those activities which involve the accumulation of money and the securing of executive power.

Aesthetic.—A high score indicates that an individual prefers and considers most worth while those activities which involve art, music, dance and literature.

Social.—A high score indicates that an individual prefers and considers most worth while those activities which involve service and help to people, and which exhibit a definite desire to respond and be with people socially.

4.—Two items per scale value from one to eleven were chosen to represent the test for each of the four values. To avoid the confusion that the negative and avoidance questions in a preliminary tryout among students received, the negative questions were eliminated. Finally there remained twenty questions per value scale or eighty items in all.

The items finally selected as they appear in the test with the median scale value and the inter-Quartile deviation are as follows:

Weighted Score	Median Scale Value	*Theoretical Items*	Inter-Quartile Deviation
10	1.00	Develop new mathematical formulas for research	(.66)
	1.50	Do research on the relation of brain waves to thinking.	(1.12)
9	2.67	Study the various methods used in scientific investigations.	(1.08)
	2.00	Develop improved procedures in a scientific experiment.	(1.00)
8	3.00	Do an experiment with the muscle and nerve of a frog.	(.87)
	3.25	Solve knotty legal problems.	(.75)
7	4.00	Make an international language.	(1.50)
	4.50	Develop new kinds of flowers in a small greenhouse.	(.87)
6	5.00	Be a scientific farmer.	(1.37)
	5.40	Do algebra problems.	(1.16)
5	6.00	Be a laboratory technician.	(1.00)
	6.00	Collect specimens of small animals for a zoo or museum.	(1.50)
4	7.00	Visit a research laboratory in which small animals are being tested in a maze.	(.80)

	7.50	Look at the displays on astronomy in an observatory exhibit	(.96)
3	8.25	Visit the fossil display at a museum.	(1.12)
	8.00	Plan the defense and offense you are to use before a tennis game.	(1.37)
2	9.00	Keep a chemical storeroom or physical laboratory in order.	(1.00)
	9.00	Read the biography of Louis Pasteur.	(1.12)
1	9.75	See moving pictures in which scientists are heroes	(.75)
	9.25	Sell scientific books.	(.83)

Weighted Score	Median Scale Value	*Economic-Political Items*	Inter-Quartile Deviation
10	1.00	Own and operate a bank.	(1.00)
	1.00	Become a U. S. Senator.	(.79)
9	2.00	Run for political office.	(.33)
	2.00	Operate a race track.	(1.25)
8	3.00	Borrow money in order to "put over" a business deal.	(1.35)
	3.00	Address a political convention.	(1.05)
7	3.60	Buy a run-down business and make it grow.	(.75)
	4.00	Be an active member of a political group.	(1.21)
6	5.20	Be a chairman of an organizing committee.	(.65)
	5.00	Plan business and commercial investments.	(1.35)
5	6.00	Lead a round-table discussion.	(1.50)
	6.00	Install improved office procedures in a big business.	(1.21)
4	7.00	Be a bank teller.	(1.37)
	7.00	Purchase supplies for a picnic.	(1.16)
3	7.67	Take a course in business English.	(1.42)
	8.00	Live in a large city rather than a small town.	(1.37)
2	9.16	Major in commercial subjects in school.	(.83)
	9.00	Work at an information desk.	(.67)
1	9.75	Collect luncheon money at the end of a school cafeteria line.	(.33)
	9.42	Be a private secretary.	(.85)

Weighted Score	Median Scale Value	*Aesthetic Items*	Inter-Quartile Deviation
10	1.33	Be a ballet dancer.	(.86)
	1.00	Paint a mural.	(.55)
9	1.66	Mould a statue, in clay.	(1.16)
	2.33	Write a new arrangement for a musical theme.	(.68)
8	2.50	Compare the treatment of a classical work as given by two fine musicians.	(1.37)
	3.00	Make a comparative study of architecture.	(1.50)
7	3.50	Participate, in a summer theatre group.	(1.03)
	3.33	Be an interior decorator.	(1.25)
6	4.50	Sketch action scenes on a drawing pad.	(1.40)
	5.00	Collect old and rare recordings.	(1.50)
5	5.75	Judge entries in a photo contest.	(1.50)
	5.75	Judge window displays in a contest.	(1.50)
4	7.00	Be a sign painter.	(1.37)
	7.25	Visit a flower show.	(1.16)
3	7.75	Plant flowers and shrubbery around a home.	(1.16)
	7.50	Make and trim household accessories like lamp shades, etc.	(1.25)
2	9.00	Listen to "jive" and "jazz" records.	(1.33)
	8.50	Dance to fast numbers.	(1.25)
1	9.60	Play the juke box.	(.66)
	9.50	Paint the kitchen white with a red border.	(.35)

Weighted Score	Median Scale Value	*Social Items*	Inter-Quartile Deviation
10	1.00	Work with labor and management to help solve their conflicts.	(.00)
	1.67	Be a medical missionary to a foreign country.	(.96)
9	1.80	Work with a group to help the unemployed.	(.73)
	1.80	Help agencies locate living places for evicted families.	(1.00)
8	3.00	Like to be with people despite their physical deformities.	(1.00)
	3.50	Treat wounds to help people get well.	(1.25)

7	4.00	Serve as a companion to an elderly person.	(.87)
	4.67	Belong to several social agencies.	(1.50)
6	5.00	Take a car load of children for an outing.	(1.50)
	5.00	Help people be comfortable when traveling.	(1.04)
5	6.33	Send a letter of condolence to a neighbor	(1.25)
	6.00	Meet new people and get acquainted with them.	(1.04)
4	7.00	Go with friends to a movie.	(1.50)
	7.50	Attend a dance.	(.83)
3	8.00	Help distribute food at a picnic.	(1.25)
	7.75	Dine with class-mates in the school cafeteria.	(1.16)
2	9.00	Play checkers with members of your family.	(1.16)
	9.00	Play checkers.	(.87)
1	10.12	Make a phone call for movie reservations.	(.83)
	9.67	Ride in a bus to San Francisco.	(.91)

In the actual test all items whose median scale-value ranged from 1.00 to 1.75 were considered as having a 1 value. The same procedure was applied to all the items so that an activity whose scale value was 6.33 was considered a 6 item, whereas a 6.89 was considered a 7 item. Moreover, for certain scoring reasons, those items having a 1 value were given a 10 weighted score, those items having a 2 value were given a 9 weighted score, etc. In short, those people who had the greatest intensity would score the most points. This reversal was convenient and in no way affected the efficiency of the scale.

RESULTS

The test or scale was tried out on 389 females and 352 males. Of the 389 females, 126 were college sophomores and 263 were high-school seniors. Of the 352 males, 121 were college sophomores, and 231 were high-school seniors. Separate norms were kept for each sex and educational group. While sex differences were found on the four value scales, no significant differences between the high-school seniors and college sophomores were obtained on any of the four scales. For all practical purposes, pooling of the high-school seniors and college sophomores scores for each sex resulted in usable general norms.

The reliability of each of the scales was computed by the split-half technique (*3*). Each scale was scored for each half of the weighted items that make up a scale. In other words, since there are two items in each scale that have a weighted value of 10, 9, 8, 7, 6, 5, 4, 3, 2, and 1, the test was scored for each half of all the weighted items and the reliability between the halves was computed. The following reliability coefficients based upon one hundred twenty-six female college sophomores were obtained: .84 for the Theoretical Scale, .82 for the Aesthetic Scale, .78 for the Economic-Political Scale and .72 for the Social Scale. The median age of this group was 18.31 with a range of 16 to 51.

The Lewerenz formula (*9*) to determine a reading grade level applied to the items, yielded a 7.8 grade level of difficulty. Despite the low 'average' reading level, however, inspection of the test indicates several words which may be regarded as above average. However, this is to be expected where the upper value intensities are to be canvassed.

One of the difficulties in using the Allport-Vernon Study of Values with the general population has been the high vocabulary grade level. Stefflre (*20*) found an 11.3 grade level for the old test. The revised edition has made an attempt to lower the vocabulary level of the test. The Lewerenz formula applied to the revised edition yielded a 10.96 grade level. This is only somewhat better.

Certain sex differences are in evidence. The median scores on each of the four scales for the two sexes are as follows: Theoretical Scale, men 45.5, women 32.0; Economic-Political Scale, men 49.0, women 32.5; Aesthetic Scale, men 38.5, women 62.0; and the Social Scale, men 56.0, women 71.5.

Further research is now being carried out to secure data on various occupations ranging from the Professional to the Unskilled group.

SUMMARY

In order to construct scales to measure the intensity of 'drive' of value activities as compared to the Allport-Vernon-Lindzey forced choice type of scale, hundreds of items were rated on a scale of intensity from lowest to highest by raters with the items of minimum variability finally selected to be used in each of four scales. Four scales resulted; namely, the Theoretical, Economic-Political, Aesthetic and Social and norms were secured on 352

males and 389 females. Fairly high reliabilities were secured. Sex differences were found on each of the four scales.

Finally it was pointed out that research is in progress to compare various occupational groups as to value intensity.

BIBLIOGRAPHY

1) H. Cantril, and G. W. Allport. "Recent applications of the Study of Values." *J. abnorm. soc. Psychol.*, 1933, 28, 259–273.

2) J. B. Cohen. "A scale for the measurement of attitude toward the aesthetic value." *J. Psychol.*, 1941, 12, 75–79.

3) Lee J. Cronbach. *Essentials of Psychological Testing.* New York: Harper & Brothers, 1949.

4) E. Duffy. "A critical review of investigations employing the Allport-Vernon Study of Values and other tests of evaluative attitude." *Psychol. Bull.*, 1940, 37, 597–612.

5) E. Duffy, and J. E. Crissey. "Evaluative attitudes as related to vocational interests and academic achievement." *J. abnorm. soc. Psychol.*, 1940, 35, 226–245.

6) S. W. Gray. "A note on the values of southern college women: white and Negro." *J. soc. Psychol.*, 1947, 25, 239–241.

7) D. Harris. "Group differences in values within a university." *J. abnorm. soc. Psychol.*, 1934, 29, 95–102.

8) E. L. Kelly, and D. W. Fiske. "The prediction of success in the V. A. training program and clinical psychology." *Amer. Psychologist*, 1950, 5, 395–406.

9) Alfred S. Lewerenz. *Word Tabulation Sheet Vocabulary Grade Placement Formula*, (copyright 1931). Los Angeles City School District Division of Instruction and Curriculum, Educational Research and Guidance.

10) C. M. Louttit, and C. G. Browne. "Psychometric instruments in psychological clinics." *J. consult. Psychol.*, 1947, 11, 49–54.

11) W. A. Lurie. "A study of Spranger's value-types by the method of factor analysis." *J. soc. Psychol.*, 1937, 8, 17–37.

12) J. B. Maller, and E. M. Glaser. *Interest-values Inventory.* New York: Bureau of Publications, Teachers College, Columbia Univ., 1939.

13) E. McGinnies. "Personal values as determinants of word association." *J. abnorm. soc. Psychol.*, 1950, 45, 28–36.

14) E. McGinnies, and W. Bowles. "Personal values as determinants of perceptual fixation." *J. Personality*, 1949, 18, 224–235.

15) T. M. Newcomb. *Personality and Social Change: Attitude Formation in a Student Community.* New York: Dryden Press, 1943.

16) L. Postman, J. S. Bruner, and E. McGinnies. "Personal values as selective factors in perception." *J. abnorm. soc. Psychol.*, 1948, 43, 142–154.

17) B. R. Schaefer. "The validity and utility of the Allport-Vernon Study of Values Test." *J. abnorm. soc. Psychol.*, 1936, 30, 419–422.

18) E. Spranger. *Types of Men.* Translated from 5th German edition of *Lebensformen* by Paul J. W. Pigors. Halle: Max Niemeyer Verlag. American agent: Stechert-Hafner, Inc., 31 East 10th St., New York 3.

19) J. C. Stanley, Jr. "Insight into one's own values." *J. educ. Psychol.* (In press.)

20) B. Stefflre. "The reading difficulty of interest inventories." *Occupations*, 1947, 26, 95–96.

21) Donald E. Super. *Appraising Vocational Fitness.* New York: Harper & Brothers, 1949.

22) P. E. Vernon, and G. W. Allport. "A test for personal values." *J. abnorm. soc. Psychol.*, 1931, 26, 231–248.

23) P. E. Vernon, G. W. Allport, and G. Lindzey. *Study of Values, A Scale for Measuring the Dominant Interests in Personality.* Revised edition, Boston: Houghton Mifflin Co., 1951.

ON THE DESIGN OF GROUPING PROBLEMS AND RELATED INTELLIGENCE TESTS

KARL MENGER

Illinois Institute of Technology

Part I of the present paper contains a critical analysis of a particular test in Figure Grouping. In Part II, a general method is developed by which correct tests of this type can be constructed. Part III contains examples which considerably widen the scope of both the critical and the positive part.

PART I. THE LOGIC OF FIGURE GROUPING

(*1*) *An example.*—An intelligence test administered during the war to a large group of college students included a question which, as far as I remember, read as follows:

> "Which of the five figures below has a property not possessed by any of the other four figures?"

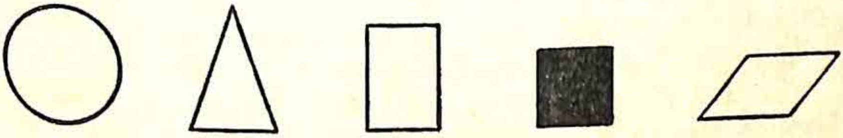


FIG. 1

Upon an inquiry, the testing agency revealed that the expected answer was 'The Square'—the square being the only one of the five figures which is black. Any other answer was held against the intelligence, more specifically, against the reasoning ability, of the tested person.

Wartime duties prevented me from pursuing the matter beyond a protest to the testing agency which had no effect. Recently, my concern was renewed when it came to my attention that questions like the one mentioned are still being used in intelligence tests.

To begin with the above example, it is clear that each of the five figures has a 'distinctive' property, that is, a property not possessed by any of the other four figures. In fact, for each figure we shall list two properties neither of which can be claimed for any of the other four figures.

(1) The circle is the only one of the five figures which (a) has no corners, (b) has infinitely many axes of symmetry.

(2) The triangle is the only one of the five figures which has (a) exactly three corners, (b) has only one axis of symmetry (namely, a vertical axis of symmetry).

(3) The rectangle is the only one of the five figures which (a) has equal angles and unequal sides, (b) has exactly two axes of symmetry (namely, a horizontal and a vertical axis of symmetry).

(4) The square is the only one of the five figures which (a) has equal sides and equal angles, (b) has four axes of symmetry (namely, a vertical, a horizontal and two diagonal axes).

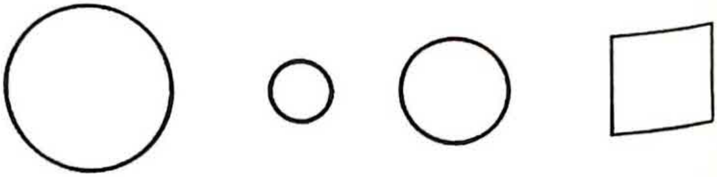


FIG. 2

(5) The parallelogram is the only one of the five figures which has (a) no vertical symmetry, (b) four angles none of which is right.

Clearly, each of the figures has further distinctive properties. For instance, the blackness of the square has not been included in the above list. Moreover, the mere property of being a circle is distinctive for the first figure; the property of being a triangle, for the second; and so on.

The situation is by no means due to a shortcoming of the particular selection of five figures. Consider, for instance, the group of three circles and one square in Fig. 2. The tested person might be expected, even more strongly than in the first example, to single out the square. But nothing is logically wrong with singling out the first circle as the only figure whose area exceeds a square inch; the second, as the only figure with an area less than one half of a square inch; the third as the only circle with an area close to one square inch.

(*2*) *The logical situation.*—The logic of the matter can be described as follows: First consider two non-identical objects. By Leibnitz' principle of the identity of indiscernibles, the non-identity of the two objects implies the existence of a property possessed by one of them, say the first, and not possessed by the other one. Clearly, the negation of this property is a property of the second object not possessed by the first.

Next consider a class of more than two objects no two of which are identical. By virtue of Leibnitz' principle, for each pair of objects belonging to the class, there is a property possessed by the first but not possessed by the second object of the pair. But the makers of tests are interested in properties of an object which are distinctive, that is, not possessed by any other object of the class. The question arises whether each object has a distinctive property (as was the case in the examples mentioned above); or whether several objects have distinctive properties while others do not; or whether none of the objects has a distinctive property; or finally whether exactly one object has a distinctive property. The presence of the last situation is implicitly assumed in the test questions. While this assumption is often erroneous, the last mentioned situation is, at any rate, the one desired by the makers of tests. We shall therefore call a class of objects a 'test class' if exactly one of the objects of the class has a distinctive property.

If the realm of properties to be taken into consideration is limited and specified, then it is easy to show that each of the four above mentioned situations can actually arise. Consider, for instance, three categories: color, form of the contour, and number of sides; and in each category two properties: white-black, dotted-solid, square-triangular. Suppose that it is explicitly stipulated that only the above categories are to be taken into consideration. Then consider the following groups of three or four objects.

Example I. A 'black' solid square, a white 'dotted' square, and a white solid 'triangle.' Each figure has a distinctive property in single quotes.

Example II. A 'black' solid 'triangle', a white 'dotted' square, and a white solid square. The first figure has two distinctive properties, the second has one, the third has none.

Example III. A black solid square, a black solid triangle, a white solid triangle, and a white solid square. None of the four objects has a distinctive property. In this example the contour of the four figures has not been varied. The example would remain valid if two contours were dotted and two left solid.

Example IV. A 'black' dotted square, a white solid triangle, a white solid square, and a white dotted triangle. Only the first object has a distinctive property, namely, blackness.

Example IV is the only one that admits a unique answer to the question "Which object has a property not possessed by the other two?" It represents the only test class. But even in this example one has to limit and specify the properties to be taken into consideration in order to be able to say that the first object is the only one with a distinctive property. The first object may cease to have this distinction

a) if properties other than those specified (such as position relative to the margin of the page or order of the objects) are taken into consideration (for instance, the second object might be characterized by the property of being immediately to the left of the vertical median line of the page);

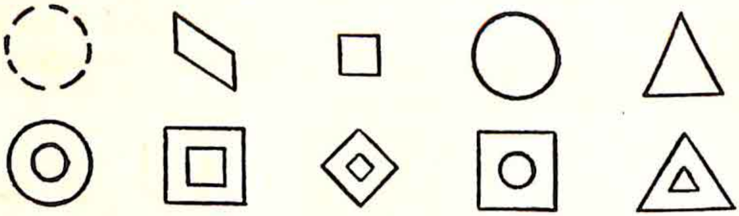


FIG. 3

b) if different specified properties (such as color and shape) are combined into one (for instance, the fourth object might be characterized as the only white square[1] in the group).

To my knowledge, however, in intelligence tests the realm of properties to be taken into consideration is not usually limited and specified.

As an empirical extension of Leibnitz' logical principle, the following law might be formulated: In every finite class of objects no two of which are identical, each object has a distinctive property.

The reader may apply the above ideas to the two examples taken from recent tests in Figure Grouping. (Fig. 3)

---

[1] As Dr. J. K. Senior pointed out, language habits have to be taken into consideration in formulating grouping tests. If in Example IV, squares and triangles are replaced by horses and cows, the German language has, indeed, a single word for "white horse" (Schimmel).

PART II. A METHOD OF CONSTRUCTING CORRECT GROUPING TESTS

(3) *The construction of test classes of five.*—If the realm of properties to be taken into consideration is limited and specified, then the construction of five objects O_1, O_2, O_3, O_4, O_5 in which exactly one object has a distinctive property is a simple combinatorial problem.

First consider five categories each containing two properties (such as white-black, dotted-solid) or, as we shall say, briefly, 'binary categories.' We shall indicate each category by a Roman numeral. In each category, we denote one property by $+$, the other by $-$. If O_1 is the only object with a distinctive property; namely, the property $+$ of category I, then O_2, O_3, O_4, and O_5 must have the property $-$ of category I. Besides, we shall let each of the objects O_2, O_3, O_4, O_5 share with O_1 the property $-$ of one category to keep them from having distinctive properties. This leads to the following scheme:

	I	II	III	IV	V
O_1:	$+$	$-$	$-$	$-$	$-$
O_2:	$-$	$-$	$+$	$+$	$+$
O_3:	$-$	$+$	$-$	$+$	$+$
O_4:	$-$	$+$	$+$	$-$	$+$
O_5:	$-$	$+$	$+$	$+$	$-$

Of course, the objects, the categories, and the properties may be permuted.

In a more economical way, a test class of five can be constructed by means of three binary categories. Again, let O be the only object with a distinctive property; namely, property $+$ of category I. The other objects shall, in addition to property $-$ of category I, have the following properties:

$$O_2\colon + +;\quad O_3\colon + -;\quad O_4\colon - +;\quad O_5\colon - -.$$

What properties O_1 has in addition to property $+$ of pair I is immaterial.

Suppose, for instance,

I is type of contour (dotted $+$, solid $-$)
II is the color (white $+$, black $-$)
III is the shape (square $+$, triangular $-$).

In the above class, O_1 is the only object with a distinctive property (dottedness). (Fig. 4)

Next consider a 'ternary' category, that is, a category containing three properties such as white-gray-black. If the three properties are distributed among five objects, then at least one object will have a distinctive property. For instance, the five objects may have the colors white, white, gray, gray, black, in which case the last object has the distinctive property of blackness. If the colors were chosen according to the scheme white, gray, black, black, black, then two objects would have distinctive properties: exactly one object would be white, and exactly one would be gray. A test class cannot contain two such objects. Moreover, it is clear that if two ternary categories are selected, then either again two (or even more) objects have distinctive properties (which is forbidden), or at least one object has two (or more) distinctive properties.

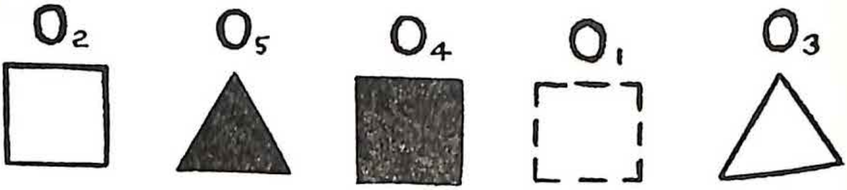


FIG. 4

We shall call a test class perfect if the only object possessing a distinctive property has only one distinctive property.

The combination of a ternary category (white-shaded-black) and a binary category (square-triangular) leads to the most economical construction of a perfect test class of five. The example, Fig. 5, includes three squares of all colors and two triangles of different colors. The square with the color not matched by a triangle is the only object with a distinctive property.

If categories containing more than three properties are introduced in a class of five objects, then the existence of exactly one object with a distinctive property is ruled out. For if four properties are distributed among five objects, then only one property occurs twice and at least three objects have distinctive properties. In the example mentioned at the beginning of this paper, one of the categories; namely shape, contained five properties (circle, triangle, rectangle, square, parallelogram). This fact alone gave every object a distinctive property.

We thus arrive at the following rule for the construction of perfect test classes of five objects with specified properties to be taken into consideration: Only binary categories (such as white-black, solid-dotted) can be used with the possible exception of one ternary category (such as white-gray-black). If a ternary category is used, then it supplies the distinctive property.

(*4*) *The construction of perfect test classes of more than five.*—Since the variety of perfect test classes of five is small, it may be of interest to see how larger perfect test classes can be constructed.

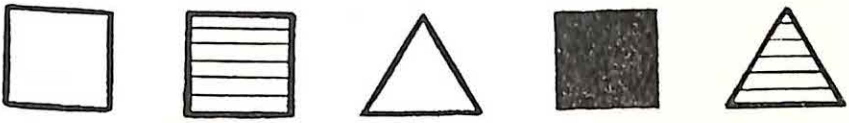


FIG. 5

Perfect test classes of six can be obtained by means of four binary categories. Below, we indicate each category by a Roman numeral, and the properties of each category by + and −.

	I	II	III	IV
O_1:	+	+	+	−
O_2:	−	+	+	+
O_3:	−	+	−	+
O_4:	−	−	+	+
O_5:	−	−	−	+
O_6:	−	−	−	−

If one ternary category is admitted, then two additional binary categories permit the construction. The three properties of the ternary category I will be denoted by +, 0, −.

	I	II	III
O_1:	+	+	−
O_2:	−	+	+
O_3:	−	−	+
O_4:	0	+	+
O_5:	0	−	+
O_6:	0	−	−

If two ternary categories are admitted, an example can be obtained as follows:

	I	II
O_1:	+	−
O_2:	−	+
O_3:	−	0
O_4:	0	0
O_5:	0	−
O_6:	0	+

Finally, it may be of interest to see test groups of eight and of ten objects which can be obtained by means of a quaternary category I containing the properties A, B, C, D, and a ternary category II containing properties a, b, c.

Eight Objects	I	II
O_1:	A	c
O_2:	B	a
O_3:	B	b
O_4:	C	a
O_5:	C	b
O_6:	D	a
O_7:	D	b
O_8:	D	c

Ten Objects	I	II
O_1:	A	a
O_2:	B	a
O_3:	B	b
O_4:	B	c
O_5:	C	a
O_6:	C	b
O_7:	C	c
O_8:	D	a
O_9:	D	b
O_{10}:	D	c

An example of the last scheme is obtained if category I denotes the number of sides of a figure,

I: 6 sides: A; 5 sides: B; 4 sides: C; 3 sides: D; and category II denotes color,

II: black: a; shaded: b; white: c.

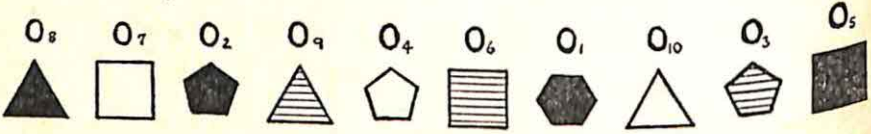


Fig. 6

The test class results are shown in Fig. 6. O_1 is the only figure with a distinctive property (having six sides).

We conclude with three examples of perfect test classes of twelve. Example 1 is based on two quaternary categories: the category I the properties of which are denoted by A, B, C, D, and II consisting of the properties a, b, c, d. Example 2 is based on the above quaternary category I, a ternary category II′ of properties a, b, c, and a binary category III′ of properties α, β. Example 3 is based on the quaternary category I and two binary categories: II″ including the properties a, b, and III″ including the properties α, β. An x in any of the examples indicates that any property of the category heading the same column may be substituted.

	Example 1		*Example 2*			*Example 3*		
	I	II	I	II′	III′	I	II″	III″
O_1:	A	x	A	x	x	A	x	x
O_2:	B	a	B	a	x	B	a	α
O_3:	B	b	B	b	x	B	a	β
O_4:	B	c	B	c	x	B	b	x
O_5:	C	a	C	a	x	C	a	α
O_6:	C	b	C	b	x	C	a	β
O_7:	C	c	C	c	α	C	b	α
O_8:	C	d	C	c	β	C	b	β
O_9:	D	a	D	a	x	D	a	α
O_{10}:	D	b	D	b	x	D	a	β
O_{11}:	D	c	D	c	α	D	b	α
O_{12}:	D	d	D	c	β	D	b	β

It goes without saying that in all these examples further categories might be added which do not supply distinctive properties. The examples given in this section attain their respective purposes with minimum numbers of categories.

PART III. APPLICATIONS AND CONCLUSIONS

(5) *Applications to other types of tests.*—It is clear that the ideas developed in both preceding parts apply to numerous types of tests other than Figure Grouping.

A recent test in Letter Grouping includes the four groups

AABC ACAD ACFH AACG

"Three of the groups are alike in some way. . . Mark the one that is different." The testee is supposed to mark the third group, ACFH, because it is the only one which does not have two A's. But, clearly, the second group, ACAD, is the only one in which the letters are not in alphabetical order, and also the only one containing the letter D; the first group, AABC, is the only one containing only the first three letters of the alphabet, and also the only one in which any two consecutive letters are neighbors in the alphabet; the fourth group, AACG, is the only one containing two consecutive letters which determine an interval of three letters between them in the alphabet (C . . . G), and also the only one containing the letter G.

In most tests in Word Grouping the situation is, practically speaking, less ambiguous although, theoretically, the differences between different types of grouping are not essential. Also with regard to words, a unique solution can be expected only if the properties to be considered are limited and specified, and a test class is constructed by the method explained in Part II. Otherwise there is the danger of a plurality of reasonable solutions due to hidden relations which can easily be overlooked in formulating the test.

Suppose a child is asked to underline the word not belonging in the group

apple, pear, strawberry, cloud, plum.

The boy underlining strawberry, because it is the only one of the five objects to which he usually has to look down, may be the potential genius in the group of tested children.

The principles developed in the preceding parts apply also to tests in Eduction of Relations. Problems of this kind are included in reasoning tests as well as in tests concerning specific knowledge. A fictitious example will illustrate the point.

van Gogh: van Beethoven = Raphael: X.

Suppose the following four choices are given for X:

a) Milton; b) Galileo; c) Mozart; d) Chaucer.

In favor of b) it might be said that Beethoven was of the same (Flamish-Dutch) extraction as van Gogh, and that only Galileo was of the same (Italian) extraction as Raphael. In favor of d) it might be said that Beethoven lived before van Gogh, and that only Chaucer lived before Raphael. c) Mozart is the only composer. By choosing Mozart one would, however, emphasize analogous properties rather than educe identical relations since it is artificial to speak of a painter-to-composer relation. One might support even the choice a) by pointing out that only Milton had an infirmity (blindness) comparable to Beethoven's deafness.

The concepts developed in the preceding sections apply also to the type of Figure Classification introduced by Spearman. In these tests, Group I differs from Group II. Each of the test symbols that belong to Group I are to be checked. In the following example, the rule which Spearman had in mind, is that the figures

Group I Group II Test Symbols

in Group I are closed, while those in Group II are open. A second rule, incidentally leading to the same classification of the four test symbols, is the existence of more than one axis of symmetry for each figure of Group I while no figure of Group II has two axes of symmetry. If the contour of a semi-circle were added as a fifth test symbol, then this symbol would belong to Group I by virtue of its closedness, but to Group II by virtue of the absence of two axes of symmetry.

If, in Spearman's example, the second test symbol is interchanged with the second figure of Group II, another ambiguity arises.

Group I Group II Test Symbols

The figures in Group I have horizontal symmetry while no figure in the new Group II is symmetric about a horizontal axis. Hence the vertical semi-circle would belong to Group II because it is open, but to Group I because it is horizontally symmetric.

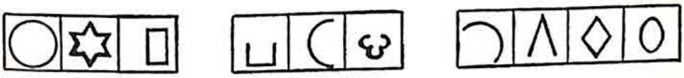


Fig. 7

These examples illustrate how the principles developed in the preceding section can be applied to the tests of the Spearman type. In this connection, a binary property is 'distinctive' if (1) it belongs to each figure of one group and (2) it does not belong to any figure of the other group. In Spearman's example, closedness is a distinctive property. Another one is the existence of more than one axis of symmetry. The property of having corners is not distinctive: in Group I, only the second and the third; in Group II, only the first and the third figures, have this property.

To construct a test of the Spearman type, two groups of figures differing in several non-distinctive and in at least one distinctive property are needed. We shall call the test 'perfect' if only one distinctive property can be found. The test is 'ambiguous' if the two groups differ in two distinctive properties and a test symbol shares one distinctive property with the figures in Group I, and the other distinctive property with the figures in Group II.

In order to avoid such ambiguities, Spearman introduced figures consisting of several components (several lines, several circles, etc.)

and used relations between these components (parallelism, proximity, equality, etc.) as distinctive properties of the figures of the two groups. In fact, in most of his excellent examples, the main difficulty for the inexperienced is to find even one distinctive property.

(6) *A comparison with number series tests.*—A number series test consists of a row of numbers which "follow one another according to some rule." The testee is expected to "find the rule and fill in the blanks to fit the rule." For instance, in 2, 4, 6, 8,—he is expected to notice that the given series consists of the first four even numbers in their natural order and to write 10, the fifth even number, in the blank.

Most mathematicians are opposed to tests of this type for the following reason: If a series of k numbers $n_1, n_2, \ldots, n_k$ is given, one can, for any number N, find a rule according to which one should write N in the blank. Hence, whatever number is written in the blank, it can be said to fit some rule.

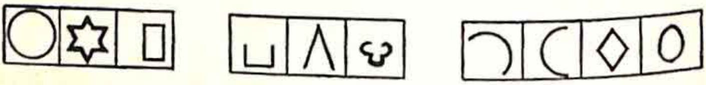


FIG. 8

Yet two important points may be advanced in support of number series test. The ability to derive, by induction, a simple rule in a short time is valuable in many domains of extra-scientific activity, as well as in science and, to a limited extent, even in mathematics. The motivation of an unexpected continuation of a given number series, as well as the actual operations leading to the unexpected number, are in most cases time consuming. For instance, to obtain 30 as the continuation of 2, 4, 8, 16 certainly takes longer than to obtain 32, and to obtain 29 would take still more time.

If, at the beginning of the test, the testees are told that they will be credited with the number of solutions attained within the allotted time, then probably few unexpected answers will be submitted. It is doubtful whether even students of mathematics would turn in many unexpected answers together with the motivating law, both obtained within the allotted time.

In contrast, in the grouping test mentioned at the beginning of this paper, it does not take longer to notice that the first figure is the only round one than that the fourth is the only black one. In grouping tests the time element is much less important than it is in the number series problems.

GENERAL CONCLUSIONS

The author believes that ambiguities of the type criticized in this paper ought to be carefully avoided in designing tests. If in the example previously mentioned the testee is expected to mark the square, then he simply is expected to guess what the designer of the test happened to have in mind. With the testee's intelligence, in general, and with his reasoning ability, in particular, the expected answer has nothing to do.

Even if there should be a high correlation between the expected answer in an ambiguous test and the positive outcome of an independent sound intelligence test, the author would not admit that an answer different from the one which happens to be expected, supplies any relevant information against the intelligence of the testee.

We should even go so far as to suggest that, on the contrary, unexpected groupings may well be a sign of superior intelligence. In order to make an answer based on a deliberate unexpected grouping distinguishable from a random answer, the testee would, of course, have to indicate the motivation or the principle of classification which has guided his choice. Unfortunately, this requirement rules out children, and makes the grading and the interpretation of the test difficult.

If a small number of more mature persons are to be subjected to an intelligence test, an ambiguous grouping test may be an excellent criterion. It would have to be phrased somewhat like this: "Find a group of four (or, if you can, several groups of four) among the following five items which have a common property not shared by the fifth. In each case state the property." Then the items would follow: words, figures, groups of letters, as the case may be.

COMMENTS ON THE CORRELATIONAL ANALYSIS REPORTED IN *INTELLIGENCE AND CULTURAL DIFFERENCES*

FRED T. TYLER

University of California

The fact of a positive correlation between socio-economic status or social class and measures of general intelligence has long been recognized. Numerous hypotheses about the nature of this relationship have been suggested:

1) Higher test scores among high-status children arise from genetic differences;

2) The environment of low-status children produces real inferiority in their intelligence;

3) Test materials are biased in favor of high-status groups;

4) Motivational patterns produce differences in test performance.

The factors indicated above are not mutually exclusive, but may operate together in various combinations.

A recent volume of Eells et al., *Intelligence and Cultural Differences*, is concerned with a number of problems in this field; only one of these will be considered in this paper: the relative bias, in different status[1] groups, of verbal and non-verbal intelligence tests.

CORRELATIONS BETWEEN STATUS AND IQ

Eells computed the correlations between the Index of Status Characteristics (ISC) and IQ's obtained from various tests for groups of over 2,000 children aged nine and ten years, and thirteen and fourteen years. The correlations for the younger group are shown in Table I (*1*, Chapter XIV).

Eells tested the significance of the difference between various pairs of uncorrected coefficients shown in Table I, reporting only one difference to be significant at the .01 level; namely, that involving the Henmon-Nelson and the Otis Alpha Nonverbal tests. He observed that the variations in the sizes of the correlations might be caused in part by differences in the reliabilities of the tests. He reported the reliability coefficients provided by the authors of the various tests and commented on the lack of comparability of these coefficients because of differences in methods of computing reliability and in ranges of talent investigated by the test makers. The reliability coefficients are shown in column 3 of Table I.

---

[1] The definitions and measures of class and status advocated by Warner and others are accepted operationally only. The writer believes that these are not satisfactory, either conceptually or statistically, to many sociologists, psychologists and educators. This does not mean that the writer is unsympathetic to or unaware of problems of individual differences connected with 'class membership.' Rather, he believes that there is need for careful definition and measurement in an area which has such important implications for educational theory and practice.

TABLE I.—CORRELATIONS AND TEST RELIABILITIES FOR YOUNGER CHILDREN

Status and IQ from	Reported Correlations	Reported Reliabilities	Corrected Correlations
Henmon-Nelson	.35	.89	.37
Otis Alpha Verbal	.34	.71	.40
Kuhlmann-Anderson	.33	—	—
Otis Alpha Nonverbal	.28	.68	.34

Eels pointed out that variations in the size of the reliability coefficients 'paralleled' variations in the size of the ISC-IQ correlations. The meaning of 'paralleled' is to be inferred from a consideration of the figures in columns 2 and 3 of Table I. However, his analysis would have been more complete if the correlations had been corrected for attenuation. The corrected coefficients are shown in column 4.[2] The difference between the correlations involving the Henmon-Nelson and the Otis Alpha Nonverbal Test is reduced, leading to some doubt about the existence of real differences between correlations of verbal and nonverbal IQ's with ISC. If status differences do arise from some bias in the test, it would appear that the bias is not exclusively that of verbal symbolism. High-status children perform better on both verbal and nonverbal tests; superior performance is more than a matter of symbolism.

---

[2] The reliability of ISC is reported by Warner to be very high. No correction was applied for unreliability in measures of status.

THE LINEARITY OF REGRESSION

When the measures of status were plotted against IQ's obtained from the various tests listed in the preceding section, Eells noticed that the regression lines were approximately linear for only part of the ISC scale. All curves tended to flatten out beyond a status

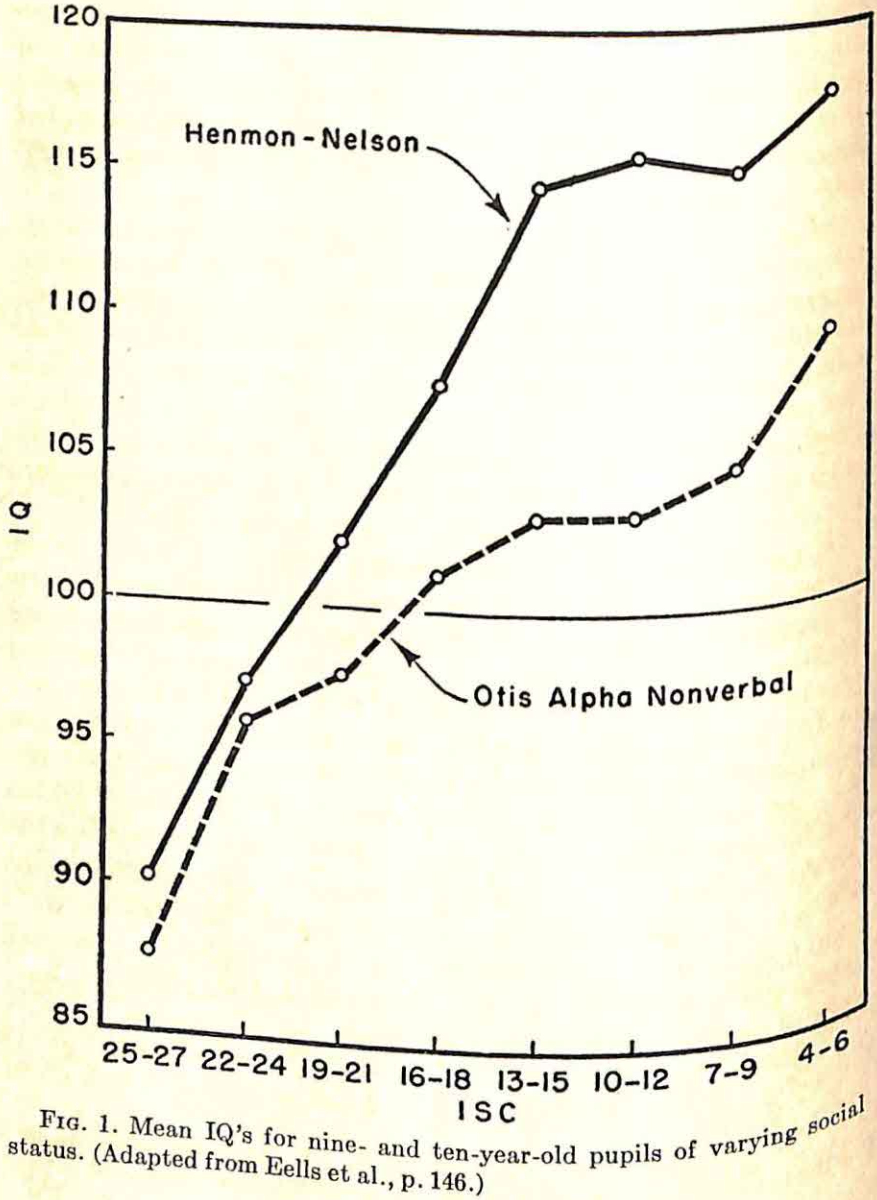


FIG. 1. Mean IQ's for nine- and ten-year-old pupils of varying social status. (Adapted from Eells et al., p. 146.)

score of 13, the lower point of what was considered to be the high-status range. He proposed several hypotheses in an attempt to account for this apparent non-linearity in the data (*1* pp. 149–151).

Two further hypotheses occurred to the writer. In the first place the levelling off might be an artifact of the measuring instruments. If tests are not of a suitable level of difficulty for the subjects, irregularities may occur at those points in the scale where the tests are too easy or too difficult.[3] Thus, if they are too easy they will not differentiate levels of ability at the upper end of the IQ scale. If they are too hard, they may also fail to be discriminatory. Unsuitable tests may provide data which do not have a linear relationship with measures of status.

Eells commented on the differences in difficulty of the various tests and of possible differences in their ceilings. For instance, several hundred children had IQ's over 130 on the Henmon-Nelson test, whereas the maximum possible IQ for a ten-year-old on the Otis Alpha Verbal test was 128, and for a nine-year-old, 136. Eells did not take these factors into account in his discussion of the levelling off found at the high-status levels. He also failed to consider the implications of the differences in mean IQ's and standard deviations of the various tests, as will be indicated later.

If a large number of subjects 'break the test,' it should not be surprising to find that there is a change in the direction of the line of regression at the point where they can no longer be measured because of insufficient 'top to the test.' The Henmon-Nelson test was designed for Grades III to VIII; the Otis Alpha for Grades I to IV; and the Kuhlmann-Anderson for Grades III to VI. The mean grade placement for the younger children in this study was 4.2. One-half of the total possible score on the Otis Alpha Verbal test corresponds to a mental age of six and one-half years; on the Nonverbal test, to a mental age of eight and one-half years. The mean chronological age of these subjects was about ten years. It is hypothesized that the lack of linearity in the data at the high levels might arise from the level of difficulty of the tests; i.e., non-linearity may be an artifact of the tests.

There is another possibility that should be considered. The IQ units for any given test may not be equal at all points of the scale. It may require more ability to raise the IQ on the Henmon-Nelson test from 115 to 120, than from 95 to 100. If this should be so, then the curve showing the relationship between status and IQ, when each is plotted as if the units were equal, cannot be expected to be linear throughout the whole range. It is generally admitted that IQ units for any given test are not necessarily equal. This possibility should, then, be considered in any attempt to account for the nature of the curves shown in Figure 1.

---

[3] A test is considered most appropriate for a group when the mean score of the group is about one-half of the maximum obtainable score.

Eells recognized the possibility that the status scale was not linear, but after some consideration discarded it as not accounting for the facts. However, until it is known that the units for the dependent and for the independent variables are equal, there is some question about the desirability of elaborating hypotheses to account for a situation which may not actually exist.

TABLE II.—MEANS AND SIGMAS ON VARIOUS INTELLIGENCE TESTS

Test	Mean	SD
Henmon-Nelson	107.2	17.2
Kuhlmann-Anderson	102.9	11.3
Otis Alpha Verbal	101.3	10.8
Otis Alpha Nonverbal	99.9	10.8

THE REGRESSION LINES INVOLVING VERBAL AND NONVERBAL IQ'S

The scattergram showing the relationship between status and IQ for different intelligence tests revealed that the mean IQ's of different status groups varied markedly from test to test. Eells noted that the means at the low-status levels were quite similar for both verbal and nonverbal tests. However, change in mean IQ with increase in status did not occur at the same 'rate' for all tests. He observed that the 'rate of increase' was greater for the Henmon-Nelson than it was for the Otis Nonverbal test. Such an observation is consistent with the third hypothesis suggested in the opening paragraph of this paper.

However, this differential 'rate of increase' of IQ with increase in status may be more apparent than real: The IQ units on different tests may not be equal. The mean and sigma values of the nine- and ten-year-olds in the study are in Table II; the IQ units are not equal from test to test.

It may seem from Table II that IQ's representing a given level of ability in this group may vary markedly from test to test. The values of IQ's, for the two tests to which Eells devotes major attention, at various sigma distances from the respective means are shown in Table III.

It is quite apparent from Table III that a given IQ does not represent the same relative ability on these two tests. A child with an IQ of about 111 on the Otis Alpha Nonverbal test stands at about the 84th percentile in this group of over 2,000 children. On the Henmon-Nelson test a child must have an IQ of about 124 to occupy the same relative position.

TABLE III.—IQ's At Comparable Sigma Positions

Sigma Units from the Mean	Henmon-Nelson	Otis Nonverbal	Difference
M − 3 SD	55.6	67.5	−11.9
M − 2 SD	72.8	78.3	−5.5
M − 1 SD	90.0	89.1	.9
M	107.2	99.9	7.3
M + 1 SD	124.4	110.7	13.7
M + 2 SD	141.6	121.5	20.1
M + 3 SD	158.8	132.3	26.5

The mean scores on the Henmon-Nelson and the Otis Alpha Nonverbal tests for each of the status groups represented in Eells' Figure 9 (p. 146) were read from the graph and changed to standard score form by using the appropriate means and sigmas. The scores thus obtained were then plotted with the result shown in Figure 2.

It now appears that the differences between the mean scores expressed in comparable units are quite uniform along the social-status scale. It is still apparent that "high-status children receive increasingly higher IQ's on all tests," but not that they do so "at a greater rate of increase for the Henmon-Nelson test." When the IQ's were changed to standard scores, the differences between the regression lines for verbal and nonverbal intelligence are much less marked than when IQ units were used for the analysis. This type of unit also indicates why it is that Eells was able to remark that at the low-status levels all tests yielded similar IQ's. It is a statistical artifact that the mean IQ's from these tests at the lowest status level shown on Figure 1 have very similar numerical values. These means are about 90 and happen to correspond to approximately the 16th percentile on both the Henmon-Nelson and the Otis Alpha Nonverbal tests.

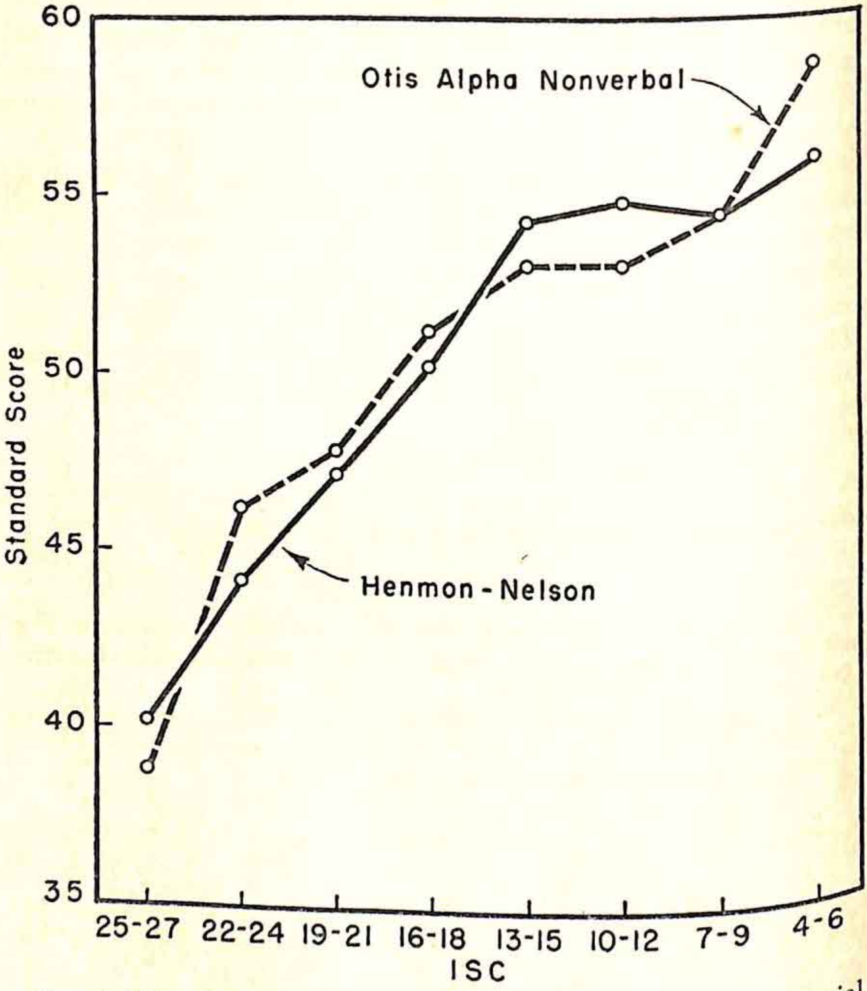


FIG. 2. Mean IQ's for nine- and ten-year-old pupils of varying social status expressed in standard score form.

Note: Standard Score = $50 + \left(\frac{X - M}{SD}\right) \times 10$

Where X is the mean score for each status interval shown in Figure 1; M is the mean of the whole group, and SD is the standard deviation, as shown in Table II.

CONCLUSION

Several features of the correlational analysis reported by Eells in *Intelligence and Cultural Differences* have been discussed in connection with the hypothesis that status differences in IQ's are a function of the nature of the tests, and that these differences are produced especially in connection with the verbal factor. A consideration of such factors as the reliability of the tests, the equality of IQ units on a given test and from test to test, and the difficulty of the tests, suggested that there is need to reconsider Eells' conclusion that "There is some definite evidence . . . that the chief reason for the status differences in IQ's may be the different opportunities which pupils from high- and low-status levels have for familiarity with the kinds of cultural materials and processes represented by the usual tests" (*1*, p. 151). Of course, the word 'may' in this quotation must not be ignored; but the materials reported in this paper indicate that the data give little positive support to even this tenuous conclusion.

It is here suggested that this type of analysis cannot provide satisfactory evidence on the factors related to status differences in IQ's, and that an experimental approach is necessary.

REFERENCES

1) Kenneth Eells, Allison Davis, Robert J. Havighurst, Virgil E. Herrick, and Ralph W. Tyler. *Intelligence and Cultural Differences*. Chicago: The University of Chicago Press, 1951, pp. xii + 388.

2) W. Lloyd Warner, Marchia Meeker, and Kenneth Eells. *Social Class in America: A Manual of Procedure for the Measurement of Social Status*. Chicago: Science Research Associates, Inc., 1949, pp. xiii + 274.

AN EXPERIMENT IN EVALUATION IN BIOLOGICAL SCIENCE[1]

JOHN M. MASON

Michigan State College

and

GEORGE W. ANGELL

State University Teachers College

New Paltz, New York

The main purposes of this study were: (1) to compare the relative effectiveness of two different evaluation programs with respect to student achievement in biological science at the end of the first term in the course; and (2) to discover if the two evaluation procedures produced any significant changes in student behavior, such as, study habits and reactions to the course.

Design of the study.—The subjects were one hundred four students enrolled in the first term of biological science at Michigan State College in the spring term, 1949. These students constituted five laboratory sections and they attended the same lectures. Laboratory sections met once each week for a two-hour period and two one-hour lectures were given each week. All the students in this study were taught in both lecture and laboratory by the same instructor.

The teaching variable was the method used for evaluating student achievement during the term. This variable was established and maintained by the design of the study whereby the students in specified laboratory sections had one evaluating program and the students in the other laboratory sections of the correlated lecture-laboratory arrangement had a different evaluation program.

In initiating the study, the instructor presented the tentative plans of the study to the students and the students in each laboratory section then made the final decision as to the evaluation program that would be used in that particular laboratory. The students in two of the laboratory sections selected one of the two evaluation programs that was deemed feasible for the study and these students were designated the experimental group. The students in the remaining three laboratory sections selected the other evaluation program and they were designated the control group. This preliminary planning, discussion, and selection of groups took place during the first three weeks of the course. The experiment proper began with the fourth week and extended through the ninth week of a ten-week term.

---

[1] Contribution No. 51 of the Department of Biological Science, The Basic College, Michigan State College, East Lansing, Michigan.

One evaluation program consisted of the administration of a weekly objective examination. Each examination was based, as a rule, upon the subject matter that had been covered the preceding week either in laboratory, in lecture, or in an assignment. Each examination was administered during the weekly laboratory period and required fifteen to thirty minutes of working time. The students who had selected this evaluation program, termed the required weekly test program, constituted the control group. There were originally seventy-four students in this group. However, the records of four of these students were incomplete and therefore were not included in this study. Thus, the control group was composed of seventy students.

The other evaluation program was arbitrarily called the self-evaluation program and the students using this program constituted the experimental group. In this program, the students were not required to take any of the weekly tests during the time of the experiment. However, the same weekly tests as were given to the control group were made available to the students in the experimental group. Students in the experimental group could take these tests either during the laboratory period or at their own convenience. A room was provided where they could secure the tests and take the tests at anytime during the school day. Keys for the tests were available and the mean scores of the control group on the various tests were also made available so that a student might compare his own achievement with that of the average of the control group.

The experimental group originally had fifty-two students, but for various reasons data were complete for only thirty-four students. Eight of the original experimental students accelerated during the term and took the Comprehensive Examination rather than the course final examination. This may indicate that the experimental group originally had a higher proportion of excellent students than did the control group. On the other hand, it may also indicate that the final average of the experimental group had been depleted by the loss of some of its best students.

It is to be pointed out that a student's course grade in the first term's work in biological science at Michigan State College was a composite grade based upon an instructor's grade for the student plus the student's grade on a departmental term-end examination. The instructor's grade constituted forty-nine per cent of the total grade and the student's grade on the term-end examination constituted fifty-one per cent of the total course grade. In this study, the instructor's grade for the students in the control group was determined by the student's scores on the weekly tests. The instructor's grade for the students in the experimental group was the student's grade on an instructor-made examination which was given to the students in the experimental group the last week of the term. Thus, the main difference between the two evaluation programs was that the students in the control group were required to take a weekly test while the students in the experimental group were not required to take weekly tests but had the test available for self-testing and self-scoring. These two programs were compared with respect to their effectiveness as shown by student achievement on a departmental term-end examination.

Hypothesis tested, collection and treatment of data, and results.—The hypothesis held for the comparison of the two groups was that achievement on the departmental term-end examination is independent of the evaluation method used during the term; that is, the mean achievement of the two groups are equal. The technique illustrated by Johnson[2] for analysis of variance and covariance with one independent variable was used to test the above hypothesis. The test of significance used was the F-test.

The independent variable held constant in the analysis was the student's score on an unannounced instructor-made test which was administered during the sixth week of the term to all the students. This test consisted of one hundred true-false items. Sixty of these items were based on the subject matter covered during the first five weeks of the term and the remaining forty questions were based on material to be studied during the last four weeks in the course. The curricular validity of this test was assumed on the fact that the items were taken directly from the syllabus used by the students. The estimated reliability coefficient of this test was .81 as determined by the Kuder-Richardson formula (approximation).[3] The respective means and sigmas for the experimental and control groups on this test were: means 61 and 59.54; and sigmas 10.14 and 11.40.

---

[2] Palmer O. Johnson, *Statistical Methods in Research*. New York: Prentice-Hall, Inc., 1949, pp. 246–255.

The departmental term-end examination was assumed to possess curricular validity in that it was constructed by a departmental committee and had been reviewed and approved by the biological science staff. The examination contained seventy-two items and the estimated reliability coefficient was .74. The respective mean scores and sigmas for the experimental and control groups were: means 43.18 and 44.00; and sigmas 7.24 and 7.58.

The results of the analysis of the data showed that the F-test was not significant. Therefore, it was inferred that the difference between the mean achievements of the students in the two groups on the departmental term-end examination could have been due to chance. In other words, it may be inferred that one method of evaluation during the term had been just as effective as the other method as far as student achievement on the departmental term-end examination was concerned.

Educational implications.—It is recognized that the preceding inference needs additional confirmation in other situations, however, this finding together with other findings points to several important educational implications. Some of these are:

1) It may be possible to set up self-evaluation procedures which are just as effective motivating influences for study and/or teaching aids as evaluation programs which require students to take weekly or frequent examinations.

2) Self-evaluation needs to be sought through instruction as does any other desirable educational objective. In this study, data were collected with respect to the use made of the weekly tests by the students in the experimental group. The per cent of students in this group that made use of these weekly tests for the six weeks of

---

[3] Henry E. Garrett, *Statistics in Psychology and Education*. New York: Longmans, Green and Co., 1947, p. 385.

the experiment beginning with the first test were 89.3, 80.8, 59.5, 46.8, 40.4 and 21.2 respectively for each weekly test. These percentages conform fairly close to the instructor's opinion as to his own efforts to motivate the students into making their own evaluations. It seems justifiable to state that the majority of college students will not assume self-responsibility for evaluation to the degree generally sought by college faculties without proper stimulation and guidance.

3) It is possible to develop self-evaluation programs whereby class time, which under the teacher testing program was required for testing, could be used for other activities.

4) A self-evaluation program, once in operation, could free teachers from many routine tasks thereby eliminating some of the time consuming operations that in themselves add nothing to teacher efficiency.

Reported changes in behavior.—In order to discover if the two evaluation procedures produced any significant changes in student behavior, an unsigned questionnaire was used to secure the data with respect to this purpose of the study. This questionnaire was administered to all students in the study during the last week of the term. The questionnaire was as follows:

BIOLOGICAL SCIENCE DEPARTMENT OF THE BASIC COLLEGE
MICHIGAN STATE COLLEGE

Basic 121

Unsigned Student Questionnaire

Instructions:

If you desire to answer *no* to the statement, mark space 1.
If you desire to answer *yes* to the statement, mark space 2.
If you feel that the statement does not apply to your section, mark space 3.
If you do not feel that you have sufficient information to give a justifiable answer to the question, mark space 4.

1) Did you have a planned study schedule which includes a time for the studying of biological science?
2) Did you keep regular hours for the studying of biological science?
3) Did you keep your preparations for biological science up to date by studying this subject at least four times a week?
4) Did you spend as much time studying for biological science each week as you did for other courses?
5) Do tests, in general, cause you to worry?

6) Do you feel that the taking of biological science weekly tests contributed to your learning of the subject matter?
7) Did you look up the answers to questions that you missed on the weekly quizzes?
8) Do you think the weekly examinations measured fairly well your knowledge and understanding of biological science?
9) Did the weekly tests make you dislike the course more than you would have without tests?
10) Do you feel, considering the effort that you have put in this course, that you have learned as much in this course as in other courses?
11) Do you feel that your work this term in biological science gave you more opportunity to find your individual study problems and to plan accordingly than do other courses?
12) Do you feel that biological science this term has provided you with an opportunity to assume more responsibility for your educational growth and development than do other courses?
13) Do you feel that your efforts in biological science have been about all that they could have been considering the other things that you have had to do this term?
14) Do you feel that if you had had more tests this term that you would have studied more on biological science?
15) Do you feel that if you had had fewer tests this term that you would have studied just about the same amount regardless of the number of tests?
16) Did the weekly quizzes cause you to cram within twenty-four hours preceding each test?
17) Do you feel that teacher evaluation is more important than self-evaluation?
18) Do you feel, considering the many factors that enter into such a feeling, that, as a general statement, you have enjoyed this course as much as other courses this term?
19) Did you enjoy studying biological science as much as you expected to at the beginning of the term?
20) Did you look upon the weekly quizzes as opportunities to determine your strengths and weaknesses?
21) Did the weekly quizzes help you plan subsequent study?
22) Do you feel that you participated in a 'democratic' teaching-learning situation this term in biological science?
23) If you had your choice of determining a testing program for the next term of biological science, would you select the method used this term?

Table I gives the percentages of 'yes' and 'no' responses to the items in the questionnaire. Critical ratios for the 'yes' responses are also given in Table I.

Items one through four in the questionnaire relate to some study habits. Inspection of Table I shows that a larger per cent of the students in the control group responded 'yes' to these items than in the experimental group. However, the percentage differences between the two groups on these items were not statistically significant. From this finding, one may infer that the requiring of weekly tests did not necessarily produce significant changes in student behavior with respect to these study habits. It is interest-

TABLE I.—PER CENT OF RESPONSES MADE BY STUDENTS IN CONTROL AND EXPERIMENTAL GROUPS TO 23 ITEMS ON THE UNSIGNED QUESTIONNAIRE

Item	Per Cent 'Yes'					Per cent 'No'		
	Exp.†	Control‡	Difference	σD_P§	$D/\sigma D_P$	Exp.	Control.	Difference
1	29.7	36.7	7.0	8.8	.80	70.2	63.2	7.0
2	25.5	32.3	6.8	8.5	.80	74.4	67.6	6.8
3	27.6	36.7	9.1	8.7	1.05	72.3	60.2	12.1
4	40.4	57.3	16.9	9.3	1.82	57.4	42.6	14.8
5	48.9	41.1	7.8	9.4	.83	46.8	58.8	12.0
6	15.2	83.5	68.3	7.5	9.11*	4.3	13.4	9.1
7	32.6	52.9	20.3	9.1	2.23*	10.8	44.1	33.3
8	13.3	67.6	54.3	7.5	7.24*	4.4	29.4	25.0
9	2.5	26.4	23.9	5.8	4.12*	20.0	73.5	53.5
10	87.2	75.0	12.2	7.1	1.72	12.7	23.5	10.8
11	48.9	26.4	22.5	9.0	2.50*	40.4	44.1	3.7
12	57.4	39.7	17.7	9.3	1.90	36.1	48.5	12.4
13	48.9	41.1	7.8	9.4	.83	51.0	57.3	6.3
14	68.0	19.4	48.6	8.3	5.85*	27.6	64.1	36.5
15	15.5	50.0	34.5	8.0	4.31*	35.5	50.0	14.5
16	4.6	51.4	46.8	9.4	4.98*	4.6	48.5	43.9
17	42.5	33.8	8.7	9.2	.95	51.0	52.9	1.9
18	78.7	72.0	6.7	8.1	.83	21.2	26.4	5.2
19	74.4	67.6	6.8	8.5	.80	25.5	27.9	2.4
20	27.6	72.0	44.4	8.4	5.29*	6.3	25.0	18.7
21	11.1	64.7	53.6	7.3	7.34*	6.6	32.3	25.7
22	84.7	69.1	15.6	7.6	2.05*	15.2	20.5	5.3
23	62.7	72.0	9.3	8.9	1.04	37.2	23.5	13.7

* Significant (Ratio 1.96 or greater).

† Calculations based on the responses of 47 students in original experimental group.

‡ Calculations based on the responses of 68 students in original control group.

§ Calculations made from formula given by Henry E. Garrett, *Statistics in Psychology and Education.* New York: Longmans, Green and Co. 1944, page 228.

ing to note, however, that 51.4 per cent of the students in the control group indicated that the weekly tests caused them to cram within twenty-four hours preceding each test (item 16). The difference between 'yes' responses for this item for the two groups was highly significant and the finding indicates that tests can be a stimulus for this kind of study.

The fact is brought out in Table I, item 5, that more than forty per cent of the students in both groups indicated that tests, in general, cause them to worry.

Since items 6, 7, 8, and 9 were marked as not applying to their section by from fifty to seventy-five per cent of the students in the experimental group no comparative interpretation of these items is offered. However, inspection of the responses of the students in the control group to these items seems to indicate that the students in the control group looked favorably upon the weekly test program. Eighty-three per cent of the students in the control group thought that the weekly tests had contributed to their learning of the subject matter and almost sixty-eight per cent thought that the tests had measured fairly well their knowledge and understanding of the course. Twenty-six per cent of the students in the control group, however, indicated that the taking of weekly tests caused them to dislike the course more than they would if they had not had the weekly tests.

The responses to items 10 through 14 give some indication of student reaction to the course. It is to be noted that the positive responses of the students in the experimental group exceeded those of the students in the control group in all of these items. The factors operating in the experimental group apparently caused a significant number of the students to feel that the procedure used in this group provided them with an opportunity to find their individual study problems and to plan accordingly than did the procedures in other courses. It is also interesting to note that the students in the experimental group felt that if they had had required tests that they would have studied more. Items 17, 18, and 19 were also answered by a larger positive per cent by the students in the experimental group than by the students in the control group, but the differences were not significant. The results also show that a larger per cent of the experimental students felt that they would not select the same program again (item 23). In general the data seem to indicate that the experiment was looked upon favorably by a majority of the students and that the students in both groups felt that they had learned as much in this course as in other courses; that they had enjoyed biological science as much as other courses; that they had participated in a 'democratic' teaching-learning situation; and that if they had a choice in determining the testing program for the next term of biological science that they would select the method used during the term.

SUMMARY

The more important findings and educational implications of this study may be summarized as follows:

1) Students, as a group, who were required to take weekly tests during the course did not score significantly higher on a departmental term-end examination than those students who were not required to take weekly tests but had the tests available for self-testing and self-scoring.

2) In terms of scores on final examinations, self-evaluation procedures can be as effective as forced evaluation procedures.

3) The two different evaluating procedures used in this study did not produce any significant changes in the study habits of the students as indicated by their responses to certain items on an unsigned questionnaire.

4) Students in the experimental group indicated that they would have studied more had they been required to take more tests.

5) More than forty per cent of the students in both groups indicated that weekly tests caused them to worry about taking tests.

6) Approximately eighty-three per cent of the students in the control group indicated that the taking of weekly tests contributed to their learning of biological science. However, approximately fifty-one per cent of this group indicated that they crammed within a twenty-four-hour period preceding each weekly test.

7) There was very little difference among students' reactions to the type of testing program that was followed in their particular group. A majority of the students in both groups reacted favorably to the experiment.

QUALIFICATION RESPONSES USED WITH PAIRED STATEMENTS TO MEASURE ATTITUDES TOWARD EDUCATION

CHARLES O. NEIDT and LYLE D. EDMISON

University of Nebraska

One of the major considerations in attitude measurement is that of demonstrating the appropriateness of the form of responding to items of a paper-and-pencil attitude scale. Inspection of published attitude scales reveals that forms of item responses are many and varied. In some cases, however, little evidence is presented to support the selection of the response form which was chosen by the test constructor for a particular measuring device.

Perhaps the factor restraining constructors of attitude scales from presenting such evidence has been the lack of independent criterion behavior with which to evaluate the effectiveness of various types of item responses. In most instances, such criterion behavior has been difficult if not impossible to identify. It seems important, therefore, that when an attitude scale is being constructed for which criterion behavior is available, several response forms for the scale should be investigated. This procedure yields evidence not only concerning the relative effectiveness of the different item response forms for the particular attitude being measured, but also evidence regarding their general applicability to other similar measurement situations in which no criterion behavior is available.

It was the purpose of this study to determine the effectiveness of using a qualification response combined with a paired statement presentation of items on an attitudes toward education scale. It was felt that such a study would be of value for determining the effectiveness of this technique as well as that of attitude measurement for predicting academic success.

PREVIOUS RESEARCH

In a previous study involving the construction of an attitudes toward education scale, Dodds (*1*) composed two hundred seventy statements describing concomitants of the educational process toward which students could reflect agreement or disagreement. When these items were presented to three hundred eighty students, she concluded that one hundred eighty items differentiated among students on the basis of total score. Neidt and Merrill (*5*) presented ninety of the most effective of Dodd's items in two different item response forms to two hundred one students. These students reacted to the ninety items in single statement form by responding on a five-point scale of the Likert type (*3*). The same students also responded to the items in paired-statement form by choosing the statement which most nearly represented their attitude. Thus, two attitude scores were obtained for each student. The score on each form of presentation was correlated with achievement, and the predictive effectiveness of each form was ascertained. It was concluded that only slight differences in effectiveness were demonstrable, but that the paired-statement form was the more feasible of the two forms since administration time for it was considerably less.

The ninety statements comprising the forty-five-item paired-statement form of the attitude scale used in this study were the same items as those used by Neidt and Merrill, except for changes in content necessitated by substituting institutional names.

THE QUALIFICATION RESPONSE

When subjects are forced to choose between two statements the one which most nearly represents their attitude, the objection is frequently voiced that neither of the paired statements presented is close enough to their attitude for them to accept it. Similarly, they often state that they would like to have chosen both statements, but were not permitted to do so. In an attempt to overcome such an objection, and to obtain further information about the reaction of subjects to each pair of statements, this study included a qualification response which immediately followed each pair of statements. The form of this qualification response was the same throughout the scale and was stated as follows:

——Only the statement I checked represents my feelings.
——Both statements represent my feelings.
——Neither statement particularly represents my feelings.

Thus, in responding to an item, the subject first checked the one of the paired statements which most closely represented his attitude, and then checked one of the three qualification statements.

For example, one of the forty-five items in complete form was:

I try to work to the utmost of my capacity in my courses.	——45——	I'm satisfied with passing grades even though I could usually do better.

——Only the statement I checked represents my feelings.
——Both statements represent my feelings.
——Neither statement particularly represents my feelings.

In summary, each subject checked one of the paired statements and one of the three qualification statements in responding to each of forty-five items.

CONSTRUCTION OF THE KEY

In constructing the keys for the paired statement response and the qualification responses, it was felt desirable to adopt a procedure which would maximize the predictive effectiveness of each item and would minimize the obviousness of the scored response. The items were presented to three hundred forty-one University of Nebraska students in the test form previously described. This group was composed of two hundred twenty-seven freshmen and one hundred fourteen seniors. Since the experimental group was to be composed of sophomores, it was felt that freshmen and seniors would yield appropriate data for the item analysis upon which to base the key.

One of the purposes of this study was to determine the effectiveness of attitudes as a predictor of scholastic success. Scholastic success was defined in this study as the average course mark obtained by each student during the semester when the scale was administered. It has repeatedly been demonstrated that scholastic aptitude scores are associated with a significant portion of the variation in course marks. To be effective as a contributor to the prediction of course marks, scores on attitude scales should also account for an additional significant amount of such variation. The criterion chosen for the construction of the key, against which to compare the responses to each item, was defined as the difference between the actual average course mark obtained by each student, and, the average course mark predicted for him from the regression of course marks on scholastic aptitude scores. This procedure tends to maximize the individual contribution of each independent variable when they are combined in a prediction scheme.

The linguistic and quantitative score of the American Council on Education Psychological Examination were the measures of scholastic aptitude used in this study. A two-variable regression equation was determined for the freshmen and for the seniors in the group. By substituting each student's L and Q score into the appropriate regression equation a predicted average course mark was obtained. The algebraic difference between the obtained average course mark and the predicted average course mark for each student was then recorded as the criterion for the construction of the key.

The test forms of the upper twenty-seven per cent and the lower twenty-seven per cent of the criterion distribution were identified and the proportions of the upper and lower group marking each one of the paired statements and each one of the three qualification statements were recorded. With the use of Flanagan's correlation table (*2*) the estimated correlation between each possible response and the criterion was obtained. Five such correlations were obtained for each of the forty-five items of the scale.

Inspection of the estimated item-criterion correlation coefficients yielded by the foregoing procedure revealed that these correlations varied from .00 to .45. A weight of 2 was assigned the paired-statement responses correlating .20 or higher with the criterion and a weight of 1 was assigned to those correlating less than .20 with the criterion. Weights of 2, 1, and 0 were assigned to the qualification statements in rank order of their correlation with the criterion.

ADMINISTRATION OF THE SCALE

The attitudes toward education scale was administered to one hundred ninety-seven students who were enrolled in a sophomore-level course in educational psychology. Of these students, one hundred forty-three were women and fifty-four were men. One hundred forty-four were sophomores, thirty-seven were juniors, and sixteen were seniors. The scale was administered during a regular class period. About twenty-five minutes was required for completing the forty-five paired statements and the qualification responses.

The average course mark each student attained during the semester in which the scale was administered was obtained from the Registrar's office at the close of the semester. The Q and L scores of the American Council on Education Psychological Examination were also obtained for each student.

RESULTS

The distributions of the two attitude scores yielded by this scale, i.e., the paired-statement scores and the qualification response scores, were plotted and found to closely approximate normality. The mean score for the qualification responses was 41.0 with a standard deviation of 4.7 and the mean score of the paired statements was 32.1 with a standard deviation of 4.9. The range for the qualification responses was from 56 to 28 or 28, and the range for the paired statements was from 43 to 19 or 24.

To obtain evidence regarding the reliability of the two scores yielded by the scale, the split-half technique was employed. To divide the scale into two forms each representing half the scale, all items were first analyzed for difficulty. The items were then paired on the basis of their percentage difficulty and one of each pair was randomly assigned to each test form. Thus two forms were obtained from the paired statements and two forms from the qualification responses. The reliability coefficients yielded from this procedure after application of the Spearman-Brown formula were found to be .68 for the paired statements and .39 for the qualification responses. The foregoing procedure of splitting the scale was followed to more nearly assure equivalence of test forms from the differentially weighted items.

To determine the effectiveness of the scale, the two attitude scale scores and the two scholastic aptitude scores were correlated with average course marks, individually and in combination. In Table I are shown the zero order, multiple, and partial correlations yielded by this procedure. Inspection of the zero order correlation coefficients in this table indicates that the qualification response score correlated nonsignificantly ($r = .053$) with the criterion and very low with the other prediction variables. The correlation coefficients between the L and Q scores and the criterion are significant but somewhat lower than have been found in other studies using comparable students at the University of Nebraska.

When the qualification response score was omitted from the multiple correlation between the criterion and the prediction variables, a reduction in the size of R from .445 to .435 was found, which corresponds to the non-significant partial correlation of .045. Thus the qualification response score did not contribute significantly to the prediction of average course marks.

When the paired-statement score was omitted from the three-variable prediction scheme, the multiple correlation was reduced from .435 to .364—a significant reduction as indicated by the partial correlation of .263 between average course marks and paired-statement scores with L and Q scores held constant.

TABLE I.—CORRELATION COEFFICIENTS FOR COURSE MARKS, ATTITUDE SCORES AND SCHOLASTIC APTITUDE SCORES

	Qualifications Response (X_1)	Paired Statements (X_2)	L Score (X_3)	Q Score (X_4)
Course marks (Y)	.053	.253	.356	.202
Qual. response (X_1)		.052	.066	.063
Paired state (X_2)			.036	.031
L score (X_3)				.479

$R_{Y(X_1X_2X_3X_4)} = .445$ $R_{Y(X_2X_3X_4)} = .435$ $R_{Y(X_3X_4)} = .364$

$r_{YX_1 \cdot X_2X_3X_4} = .045$ $r_{YX_2 \cdot X_3X_4} = .263$

IMPLICATIONS

The results of this study provide evidence for two inferences. First, although the use of the qualification response combined with the paired statement response overcame the objections frequently voiced by subjects when they are forced to choose between paired statements, its quantitative score was not effective under the circumstances of this study. Perhaps other testing circumstances and scoring procedures will yield more favorable evidence.

Second, the significant contribution of the paired-statement score to the prediction of academic success emphasizes the possibilities of combining attitude measures with other prediction variables for obtaining more accurate prediction. Although the forty-five items used in this scale had been rather carefully analyzed and pretested before this study was undertaken, they undoubtedly can be improved and additional effective items can be constructed. Evidence from this study supports the position that attitude measurement can significantly improve the prediction of academic success.

SUMMARY

An attitudes toward education scale composed of forty-five paired statements each followed by a three-category qualification response was administered to one hundred ninety-seven University of Nebraska students. The key for scoring this scale was constructed by obtaining the estimated correlation between each possible item response and an academic achievement criterion in another sample of three hundred forty-one students. The criterion for the key construction was obtained by subtracting an expected average course mark (predicted from each student's scholastic aptitude scores) from the average course mark which each student actually attained. Weighted scores were obtained for the paired statements and for the qualification responses of the one hundred ninety-seven students.

Partial correlation coefficients were obtained between average course marks and the two attitudes scores with L and Q scores of the ACE held constant. These correlations were found to be .045 for the qualification response scores and .263 for the paired statement scores. It was concluded that the score on the qualification response did not contribute significantly to the prediction of course marks, whereas the paired statement score yielded a significant contribution to the prediction scheme.

SELECTED REFERENCES

1) Barbara Degen Dodds. *Relative Effectiveness of Paired Statements and Single Statements in a Scale Designed to Measure Attitudes toward Education*. Unpublished M. S. Thesis. Ames, Iowa, Iowa State College Library, 1949.

2) John C. Flanagan. "General considerations on the selection of test items and a short method of estimating the product-moment coefficient from data at the tails of the distribution." *Journal of Educational Psychology*, 30: 674–680, 1939.

3) Rensis Likert. "A technique for the measurement of attitude." *Archives of Psychology*. No. 140, 1932.

4) R. C. Meyers and D. G. Schultz. "Predicting academic achievement with a new attitude-interest questionnaire." *Educational and Psychological Measurement*, 10: 654–663, 1950.

5) Charles O. Neidt and William R. Merrill. "Relative effectiveness of two types of response to items of a scale on attitudes toward education." *Journal of Educational Psychology*, 42: 432–436, 1951.

6) William A. Owens. "Item forms and 'false-positive' response on a neurotic inventory." *Journal of Clinical Psychology*. 3-264-269, 1947.

BOOK REVIEWS

Henry Bowers. *Research in the Training of Teachers.* Toronto, Canada: J. M. Dent & Sons (Canada) Ltd., 1952, pp. 167. $1.90.

From the standpoint of the lay reader's understanding of reports of scientific investigations this little volume could well serve as a model. Since most of those concerned with teacher-training are not generally highly trained in statistics, this ease of reading is a most commendable characteristic. Nor does this imply any evidence that the book's author is not thoroughly conversant with the statistical tools appropriate to his task. He evidently has kept always in mind the reading audience.

The book represents an enormous amount of patient, well-directed, panistaking labor. The spirit in which the book was written is well stated in the Preface: "Only a small part of the research carried out in the Stratford and Ottawa Normal Schools is included in the papers now published; nine-tenths of the glacial mass is submerged. This submergence was the result of self-criticism which, at least, felt ruthless, and a desire to maintain inner standards of procedure. These remarks are made without implication that tables have been handed down from Mount Sinai."

The contents are organized as fourteen 'Papers,' each of which reports one or more experiments to determine the importance of a rather exhaustive list of variables: "I. Condensation of the Academic Records of Ontario Secondary School Pupils Who Possess the Academic Requirements for Entering a Provincial Normal School; II. The Appearance of the Student-teacher; III. Concomitants of the Marks Obtained in the Term Examinations of the Normal School; IV. Traits of Personality Associated with the Degree of Success of Student-teachers in Practice-Training; V. Concomitants of the Marks Given for Practice-teaching; VI. The Bearing of Miscellaneous Activities and Interests of Student-teachers on Achievement in Practice-teaching; VII. Qualities of Sociability and Leadership Possessed by Student-teachers; VIII. The Relationship of Height and Weight of Student-teachers to the Quality of Their Practice-teaching; IX. The Aptitude Test for

Elementary School Teachers-in-Training as an Instrument for Prediction; X. Growth of Student-teachers in Certain Traits of Personality During Their Period of Training; XI. Sex-differences in Ratings of Certain Traits of Personality; XII. A Comparison of the Achievement of Male and Female Student-teachers; XIII. Variation of Standards in the Marking of Lessons Taught by Student-teachers; XIV. The Feasibility of 'Homogeneous' Grouping in a Normal School."

Criterion measures were in general such as are assumed to be related to effectiveness in teaching. It is to be hoped that these well-designed, careful researches will be extended to include the more ultimate criterion of degree of desirable changes in pupils taught by teachers with differing characteristics. A readily available criterion not investigated is obviously the attitude of pupils taught by student-teachers, a variable almost certainly closely related to pupil motivation and learning.

A few rather minor criticisms include the following: There is no index—not too serious, however, since each paper is relatively short and self-contained with a summary and conclusions. Tables are not numbered in sequence throughout the book, but within each paper. The one graph (p. 3) rather confusingly reverses the usual Cartesian coördinate system on the abscissa. The reviewer detected a few errors, either typographical or computational. For example, the Spearman-Brown *r* on page 12 is incorrect; it should be .84, not .88. The first percentile value given in Table I, p. 14, must be in error. A page reference on page 27 is given as 'page ■■■'. On page 59 the two rho's should be .74 and .73, respectively, not .60 and .78. Some account might well have been taken of the extensive—albeit not too productive—research literature on teacher effectiveness in the United States.

These criticisms, however, are minor as compared to the light thrown on the import of student-teacher characteristics and on training procedures in the normal school situation in Ontario. Few teacher-training institutions are fortunate enough to have as chief administrator such a scientifically-minded and pertinacious pursuer of facts and generalizations relevant to his highly important professional job.

H. H. Remmers

Purdue University

OTTO POLLACK ET AL. *Social Science and Psychotherapy for Children.* New York: Russell Sage Foundation, 1952. pp. 242

This is the result of a joint effort made possible by the Russell Sage Foundation and the Jewish Board of Guardians and directed by Otto Pollack with a group of nine collaborators. It attempts to give evidence, with considerable success, of "significant benefits derived from the integration of pertinent data and principles selected from the fields of sociology, cultural anthropology, and social and learning psychology into the psychoanalytically oriented child guidance practice of that agency." (p. 7.) "This book is a report of exploration into the question of whether existing funds of social science knowledge can be adapted to psychotherapy practiced in a child guidance setting." (p. 9.) It presents the therapeutic approach of the Jewish Board of Guardians in the belief that it will have interest and applications for a wide audience.

The subjects dealt with in various chapters by different collaborators include: adapting social science to child guidance practice; the concept of 'Family of Orientation' in diagnosis and therapy; social interaction and therapy; extra-familial influences in pathogenesis; culture and culture conflicts in psychotherapy; age-sex rôles and psychotherapy of adolescents; the therapeutic management of anxiety in children; the utilization of volunteers in sociodynamic psychotherapy; limited treatment goals; and an evaluation from the psychiatric point of view.

The results presented in this volume are those of a two-year venture to develop a "liaison between the behavior sciences and a specialty in social practice, child guidance." (p. 7)

One of the special emphases is upon 'The Concept of Family of Orientation.' This means "the sum total of persons who form continuing members of the household in which the child grows up, that is, the primary group at the home." (p. 42). This must be carefully distinguished from the 'family of procreation.' Failure to make this distinction is thought to be very serious and to occur altogether too frequently.

Interpersonal relations, of course, are of prime concern throughout the several discussions and the value of the volume lies not only in the general overview of these relationships, but also in the pertinent and concrete illustrations of the kinds of difficulties that arise in individual cases. These should be known and understood by the clinical worker but often are not known.

Drinking milk during meals is a violation of the food ritual of Jewish law. The worker did not understand this. The culture of the worker did not permit understanding of the particular problem of the child. Some customs in camp are not what the child has been taught is 'kosher.' Situations arise which the non-Jewish worker may not understand. "Camp experiences may require the child to eat food which is not prepared according to the orthodox food ritual practiced in the home." In general, cultural differences between worker and patient may lead to serious misunderstandings and uninformed methods of therapy.

The worker needs to understand that one may be quite normal and adjusted to situations that follow his education and training but seriously disturbed in an environment in which his habits and customs are neither followed nor understood. An intake worker recorded a "referral from a Hebrew school, apparently not being aware of or not appreciating the difference in ideology between a socialistically oriented Yiddish school and the orthodoxy and conservatism of a tradional Hebrew school (Cheder)." (p. 122).

The report explains what seems to have been a successful coöperative effort in the development of means and methods in psychotherapy. The point of view is essentially psychoanalytic and the concern of the authors is with "the particular therapeutic approach adopted by the Jewish Board of Guardians." The concrete examples make it especially of interest and value to all students of psychotherapy.

A useful and well made index (pp. 235–242) concludes the volume.

A. S. Edwards

The University of Georgia

G. Milton Smith. *More Power to Your Mind.* New York: Harper and Brothers, 1952, pp. 180.

This popular treatise on mental hygiene is concerned with the daily problems of life and of personal relationships. It is designed for those individuals who are needlessly operating at a level of effectiveness which is unnecessarily low due to relatively minor conflicts and frustrations. Self-understanding is fostered so that one may deal intelligently with everyday problems of personal, business, social and family life. A one-sided approach is avoided. The author believes that "no system which focuses exclusively on training the mind, or disciplining the body, or exalting the spirit is adequate to bring our performance in line with our capabilities. We need all three."

In the thirteen short chapters, discussions are concerned with such topics as needs of the self, how emotional needs and conflicts arise and how they may be dealt with, learning for effective living, satisfaction derived from work, the mind-body team, adjustments to sex needs, and the rôle of family life in effective living.

Although written in popular style, this book is based upon sound psychological information. It is written in a simple, clear and forceful style. The balance and sense of proportion are excellent. Since the discussion presupposes no technical background, the material will be readily comprehended by any intelligent reader. This book should not only have great vogue, but also should be a boon in promoting better adjustment through an improved understanding of self needs.

MILES A. TINKER

University of Minnesota

MARGARET MEAD AND FRANCES COOKE MACGREGOR. *Growth and Culture: A Photographic Study of Balinese Childhood.* New York: G. P. Putnam's Sons, 1951, pp. 223.

The work of Arnol Gesell and his colleagues has made a significant and lasting impression on the science of developmental human behavior. Their findings concerning specific growth patterns in infant and child development, and their concept of 'growth spirals,' have influenced profoundly the thinking of present-day child psychologists, educators, and many parents.

In this book, two well-known anthropologists have attempted to see to what extent the various growth patterns which the Gesell workers found in their New Haven subjects would appear in children living in a village of Bali (now a part of the Republic of Indonesia). They also were interested in finding out whether or not the Balinese culture would cause significant differences to appear in these children as contrasted with Americans. What they discovered was this: The general stages of developmental behavior are very much the same for Balinese as for American children.

There are some differences, however, the sequence from creeping to walking being an outstanding one. Seen against Gesell's spiral analysis of development, the Balinese children tend to neglect the crawling stage, and to some degree the creeping stage as well, but they reinforce the frogging-squatting sequence that American children neglect. Other developmental areas in which contrasts appear are related primarily to the emphasis on flexibility and high tonus which is encouraged in Balinese children.

Emerging from these and other comparisons of developmental behavior is the keynote of the entire book: maturation and learning interact in the production of the various developmental stages studied. While biology so fixes these stages that in a great many ways Balinese development is like that of American children, yet cultural differences also go to determine differences in developmental patterns.

The reader whose background is closely related to cultural anthropology will probably find to his satisfaction that this already widely-accepted principle of maturation-learning interaction is ably discussed and amply buttressed with appropriate observational data and interpretations. In addition, he may well find, along with the reader whose primary training and interest lies somewhat outside of anthropology, that the most rewarding part of the book is the set of fifty-eight photographs of Balinese children and adults which illustrate the character of the various developmental stages and relationships described in the text. (These plates were selected from pictures made by Gregory Bateson between 1936 and 1939, when with Margaret Mead, he collected data for his later published work in 1942, *Balinese Character*.) The arrangement of the photographic plates, the accompanying explanations for each, their relevance to the developmental tasks discussed, and the sheer photographic fidelity with which they present what the authors are trying to explain—these are all excellent, and, in the reviewer's opinion, are what really make the book different and new.

These photographs drive home, indeed, the principle that culture can modify developmental behavior patterns. But they also show—and this is a virtue of the photographic approach, that it can do this so dramatically and clearly—how very much like their American age-mates, after all, these Balinese children are.

University of Illinois FRANK COSTIN

JAMES L. MURSELL. *Using Your Mind Effectively*. New York: McGraw-Hill Book Company, 1951, pp. 254.

In spite of the more general implication of the title and the author's protestation against confining the concept of mental effectiveness to scholastic efficiency, *Using Your Mind Effectively* can be described best as a 'how-to-study' book written in popular style. The theme upheld throughout the volume is that the achievement of better thinking is the requisite for attaining greater adeptness in studying. Since the desideratum of better thinking is learnable and teachable, Mursell's task is to show the way to the reader. The steps in reaching comprehension and understanding are formulated as: (1) obtaining an over-all view of the material under consideration, (2) identifying its main features, and (3) working the details into their proper relationship to the main skein.

After an introductory section on the value of mental efficiency, the book is broken down into three main divisions; one dealing with an exposition of the general psychological principles to be observed in attaining intellectual mastery and two devoted to the demonstration of practical applications. In Part One the author elaborates on the nature and usefulness of the sequence; picking up the main thread, identifying essentials, and relating the details to the whole. He discusses the importance of this sequence in the reading of textbooks, study of college courses, and the writing of term papers and theses. The construction of a mental map into which essential parts and details may be fixed is offered as the *modus operandi* for achieving mental excellence. Applications of this technique to non-academic situations are demonstrated, e.g., preparing a talk, orienting oneself to a new job, arranging one's budget, and learning to play tennis.

In Part Two standard topics of the 'how-to-study' books such as budgeting time, note-taking and note-using, self-testing, and concentration are discussed under the rubric of working tools and practical plans. The general rationale for the recommended approaches as well as the minutiae of procedural detail is set forth. Much of this is the stock advice on these topics of 'how-to-study' books. The author appears at times to be self-conscious about some of the trivia which he presents. After marshalling a lengthy list of materials and equipment that should be kept in the study room, he feels constrained to remark that he offers this for whatever it is worth. In this division of the book the *leitmotif* is that the facilitation of thinking should be the governing consideration in the arrangement of time, conditions, and methods of work. Everything should be instrumental to the goal of encouraging thinking; note-taking and review are seen as experiences in thinking.

In Part Three under the title "Some More Extended Applications" appear the following topics: memorizing, reading, writing term papers and theses, and creative thinking. Good memorization is viewed as conditional upon proper purposing, understanding, and noticing. Analyzing, interpreting, and thinking are considered the crux of the development of reading skill. In writing term papers, it is recommended that the outline and over-all plan be completed before a single word of the final manuscript is written. In a very short chapter on creative thinking, Mursell makes the point that whenever one grasps the meaning of another person's writing, one is engaged in creative thinking. The author makes clear that his aim in offering suggestions is not to facilitate the functioning of routine-minded workers, but to influence the intellectual worker to approach even mundane tasks as invitations to creative thinking. He concludes his work with the statement that he has tried to make everything in the book center on creative thinking.

After finishing this book, the reader might legitimately ask what progress psychology has made, since William James, in offering assistance to students in their pursuit of academic success. No later writer seems able to go beyond or even approach the level of James' *Talks to Teachers on Psychology: and to Students on Some of Life's Ideals* published in 1899. Basically little that is new since Kitson's *How To Use Your Mind* appears evident in the more recent books on 'how to study'. Aside from the Q 3 R technique of Ohio State University for studying chapters in the conventionally organized textbook and the 'push on' or quick reading method for mastering foreign languages, there is in the book under review scarcely any reference to new developments in effective study methods. Psychologists who choose to write 'how-to-study' books seem to feel obliged to approach their topics as exclusively cognitive-intellective matters unrelated to the emotions and the total personality. They appear to be oblivious to the psychoanalytic movement and to dynamic psychology in general. Mursell fails to suggest that the inability to understand and think effectively may be due to emotional factors as much as to poor techniques in cognition. The failure to integrate cognitive functions with the entirety of the personality accounts for the antediluvian character of most 'how-to-study' books.

Mursell has done a good job of expounding the traditional approach for developing mental efficiency. However, at times, the style is repetitious, long-winded, and too general without the compensation of being inspirational—hence boring. The standard of writing at times seems to dip below the college level to which it is apparently directed. The attempt to offer applications to non-academic pursuits appears to be simply a gesture. For those who want a simple and very clear exposition of the conventional views on effective use of the mind, this book may be most satisfying.

PHILIP M. KITAY

Adelphi College

C. W. ODELL *How to Improve Classroom Testing*. Dubuque, Iowa: William C. Brown Company, 1953, pp. 156.

Professor Odell has prepared a student manual that emphasizes the practical and non-technical aspects of the construction and administration of informal or teacher-made tests of achievement. Students of education will be pleased to find that he sets the problem of measuring achievement in the context of curriculum development by taking the objectives of education as the definitions of the achievements that are desired. He then concentrates on testing programs and types of tests, excluding problems of intelligence and personality measurement. Chapters IV through XII are devoted to test construction principles and illustrations of various types of test items or problems; these illustrations are drawn from a number of subject matter fields. The manual appears to be a good source of ideas for test item types, and as such it should be of value to teachers. One chapter on administration and scoring and one on statistical methods in connection with testing complete the volume. The chief contribution of the manual is its rather extensive and practical advice on how to develop and use informal tests.

CHESTER W. HARRIS

The University of Wisconsin

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 44 October, 1953 Number 6

CONCEPTS OF ORGANISMIC GROWTH: A CRITIQUE

FRED T. TYLER

University of California

The educational programs of American public schools have been based upon a variety of theories respecting content and methodology. At one time religious beliefs and values played a prominent rôle in determining the selection of materials and procedures. More recently, other bases have been proposed, as, for example, the social utility theory, which selects materials in terms of practical values and eliminates topics regarded as not related to adult needs.

During the past two decades greater emphasis has been assigned to the needs of the child himself, and to the facts, principles and generalizations of growth and development as curricular guides. There is every reason to believe that an accurate knowledge of growth and development provides one very real and meaningful foundation for the curriculum of both elementary and secondary schools. Psychologists and educators must, however, be concerned about the accuracy of their data and the validity of proposed implications. It is the purpose of this paper to consider some of the principles of growth which are discussed by two writers (W. C. Olson and C. V. Millard) as having educational significance, to scrutinize the supposed function of these principles, and to study their suggested educational implications.

In advocating a modified basis for instructional programs and materials, Millard suggests that the word 'growth' may eventually replace the term 'learning'. Such being the case, he argues for a clear understanding of 'growth' in order to avoid distorted generalizations and implications for classroom procedures (*8*, p. 9). He further contends that "Growth as used here may justifiably refer either to learning which is permanent and meaningful or to given biological change" (p. 10); and that "Growth in school subjects is governed by the same laws that govern physical or mental growth" (p. 11).

Olson's writings, too, stress the importance of the principles of growth and development as guides for the curriculum, including instructional methodology: "As will be seen in later discussions of the stimulation and retardation of growth, the individual himself and the adults about him have little control over physiological time; wise nurture, therefore, in both physical care and education, makes no attempt to alter the individual's rate of growth and development" (*13*, p. 17).

A few more quotations from Olson and Millard will indicate further some of their ideas concerning the implications of data on growth and development for the organization of educational programs: "Enough is known . . . to question seriously remedial work, drill, and other teaching devices" (*8*, p. 51). "The most recent findings on growth interrelations tend to challenge much of the work now classified as studies of readiness" (*8*, p. 53). "The writer believes that the late and abrupt upswing in the growth of the testes in boys is more than coincidental with similar changes in reading" (*13*, p. 133).

Some of the concepts and implications cited above are open to serious question. The material to be discussed in this paper includes: (a) The concept of the 'growth age' and the 'organismic age'; (b) Patterns of growth; (c) Physical growth and reading readiness; (d) The cyclical nature of physical growth and educational progress; (e) Familial resemblances; (f) Displacement, convergence and perturbations.

THE CONCEPTS OF GROWTH AGE AND ORGANISMIC AGE

1) The concept of the 'growth age' as the unit for measuring 'growth' in both structures and functions has been developed most extensively by Willard Olson. He explains that the principle involved in the determination of growth ages is the same as that which has been employed in the preparation of mental age scales. He further states, "that the age unit method translates unequal units of the original measure into equal units of age" (*11*, p. 3).

This is a rather unusual assertion. In the psychological literature on mental measurement mental age units are not usually considered to be equal. Richardson, for example, observes that the mental "... age scale meets not one of the three requirements set up (for a measuring device); it has no real origin of measurement; its units are not equal, and it does not isolate a single, unitary variable for measurement" (*14*, p. 26).

Terman, the foremost exponent of the mental age unit for measuring intelligence, has never claimed that the units are equal (*16*). Similarly, McNemar has pointed out: "Not only have no claims been made for the equality of mental age units but actually their inequality has been admitted (see pages 24–29 of *Measuring Intelligence*). In fact, the use of IQ units is predicated on inequality of mental age units" (*10*, p. 153). Recognition of the inequality of these units does not mean that age scales are without value, but it does place certain limitations upon their use.

These remarks about mental age units surely apply to other age units which are based on the same conceptions and derived in a similar manner. On theoretical grounds, there seems little justification for believing that the difference between Reading Ages (RA's) of nine and eight years represents the same psychological difference, i.e., difference in achievement, as that between RA's of thirteen and twelve years, or the same structural difference as that between Height Ages (HA's) of thirteen and twelve years.

2) Olson believes that changing growth data into growth ages enables one to "refer such diverse things, for example, as height in inches, weight in pounds, number of teeth erupted and strength of grip in kilograms to a common growth age scale and speak of height ages, weight ages, dental ages, and grip ages" (*11*, p. 2). In another publication he states that, "The unconventional liberty taken in averaging diverse data is based on the hypothesis that all are samples of structures and functions of the organism, and that different individuals may be organismically equal while having different patterns" (*1*, p. 202). The average referred to in this quotation is designated by Olson as the Organismic Age (OA).

The use of growth ages to determine the OA will be illustrated by an example from the manual prepared by Olson and Hughes (*11*, p. 14, Exercise 3). The following data for a twelve-year-old boy were obtained from their tables:

Characteristic	Absolute Measure	Growth Age in Months
Height	58.7 inches	152
Weight	106.0 pounds	176
Dentition	17 teeth	127
Intelligence		136
Education		140
Sum of ages		731
Mean of ages		146

The arithmetic mean of the five ages, 146 months, is this boy's organismic age.

It is a cardinal principle of measurement that the units of various scales must be equal if the results of measurement are to be added, as is done in computing the OA. As has been suggested, there is some doubt that growth age units are equal. An age scale may provide useful data for portraying a specific aspect of growth, but it is doubtful if data from a series of age scales can be appropriately used to make a direct comparison of several types of 'growth' in a given child, or to compute such statistics as the OA.

Despite Olson's assertion to the contrary, one must also ask about the psychological meaning of an average computed from several different measures of an individual, even when they are all obtained from a single type of measuring instrument, such as a linear scale. Consider the following data for two subjects in the Adolescent Study at the Institute of Child Welfare, University of California:[1]

Subject	Growth Age in Years		Mean Growth Age
	Height age	Biacromial age	
A	12.5	10.5	11.5
B	11.2	12.0	11.6

Here are two boys with very similar mean growth ages. In what meaningful sense can it be claimed that they are 'organismically equal'? Knowing that each has a mean growth age of about 11.5 years gives no useful clue about their comparative status in height and shoulder width, the characteristics contributing to this age. Surely one may question the significance or utility of an average based upon these measures.

---

[1] Growth ages are based on norms from this particular study.

It seems even more difficult to interpret the arithmetic mean of a series of ages, some of which are measures of size and some of other aspects of structural development, such as, dentition and ossification. Finally, is the OA really a useful concept when it is the mean of a series of ages representing status in height, strength of grip, basal metabolism and spelling achievement—measures of both structure and function?

What is the educational or clinical value of knowing the organismic age of the following Adolescent Study boys?

Subject	Wt. (Kg.)	Wt. Age	Ht. (mm.)	Ht. Age	MA	IQ	CA	OA
D	347	10.8	1509	12.4	12.3	102	12.1	11.8
E	391	12.0	1476	11.8	11.7	105	11.1	11.8
F	338	10.5	1480	12.0	13.1	114	11.5	11.9

D is a relatively tall, slender adolescent. E, one year younger, is shorter but heavier. Based on weight age, height age and MA, their organismic ages are the same.

F is somewhat brighter than E but is considerably lighter. They differ in both MA and WA. Despite the comparability of the OA's, these boys may be expected to have different kinds of problems in their adjustment to school and to their peers. Knowledge that they are organismically equal (that is, have equal OA's) seems less valuable than the knowledge of their specific status in height, weight and intellect.

PATTERNS OF GROWTH

Olson states, "Whole classes and school systems have been shown to have distinctive patterns depending upon selective factors. Thus, children selected for mental deficiency tend to have a low MA curve and to regress toward the average on other factors. Children selected for intellectual giftedness tend to have a high MA line and to regress toward the mean on other factors" (*13*, p. 182). Two points should be noted about the interpretations included in this quotation. It is not surprising to find that a group of children selected for intellectual deficiency regress toward the mean of unselected children on other measures of 'growth'. Neither is it surprising to find that children selected for 'intellectual giftedness' tend to have a high MA line. 'Intellectual giftedness' is defined by a superior mental age. The regressions and tendencies which attract Olson's attention are statistical artifacts. They cannot be used as evidence of some underlying principle of 'growth', nor as evidence for distinctive types of individuals.

Millard states that "... It is impossible to differentiate one growth pattern from another when several are reduced to growth ages and graphed together. Stated in another way a reading curve and a height curve for the same child are usually strikingly similar" (*8*, p. 22). A consideration of the growth curves, and the data for such curves, presented by Olson and Hughes in their *Manual* (*11*) warrants some suspicion concerning the validity of this conclusion.

The data for the child presented in Figure 1 of this Manual (*11*) were used to test the hypothesis that it is impossible to differentiate one growth pattern from another when they are both expressed in age units. On the assumption that growth curves (in age units) are linear, the equations for the curves of Educational Age (EA) and Dental Age (DA) were computed. If the growth curves of different growth phenomena are 'strikingly' similar, then the equations should reveal this in the similarity of their slopes and intercepts. The slopes and the intercepts of these equations are not identical, as may be seen from the equations (1) and (2) below. The values of the DA's were read from Figure 1; the EA's were taken from the table on page 7 (*11*). Letting y be the dependent variables (EA and DA) and x the independent variable (chronological age—CA) the equations for the curves showing age changes for EA and DA, respectively, were computed by the writer and found to be:

$$(1) \quad y = 1.879x - 64.814$$

and

$$(2) \quad y = .776\,x + 23.036$$

If the curves are such that is is not possible to differentiate one type of growth from another, then it should be possible to use either equation to predict EA or DA. Using the two equations in turn, the predicted EA's at 122 months are about 164 and 118 months, a difference in predicted value of nearly four years; measured EA was reported to be 174 months. DA, then, is an unsatisfactory predictor of EA. The presence of poor predictions cannot be ignored when a generalization is being attempted; exceptions as well as agreements must be considered.

In conclusion, it seems that more evidence must be forthcoming before the term 'growth' can be applied to both structural and functional changes as is proposed by Millard: "Growth as used here may justifiably refer either to learning which is permanent and meaningful or to given biological change" (*8*, p. 10).

MENTAL AND PHYSICAL QUOTIENTS

Olson and Hughes suggested the Organismic Quotient (OQ) as a measure of rate of organismic growth (*12*). They defined the OQ as the ratio of organismic age to chronological age; thus it is comparable to the Intelligence Quotient (MA/CA). In their paper, they presented the OQ's for a child at ages five through eleven, these being, 101, 102, 98, 99, 98, 102, and 106. The greatest difference is eight OQ points between ages seven or nine and eleven years of age. Olson commented on the stability represented by the OQ. These quotients may indicate that the OQ is very stable, more so than is the IQ. However, studies comparable to those reported by Honzik, Macfarlane and Allen (*6*), and by Bayley (*3*) would be necessary before such a generalization should be proposed. Furthermore, greater stability should be found since the OQ, being a mean of several ages, can be expected on statistical grounds to be more stable than the IQ.

Millard was interested in the possibility of predicting IQ's from measures of height maturity (*7*). Using the Courtis method of determining the equation expressing the relationship between age and measures of growth, he predicted the IQ's of twenty-five boys from a knowledge of their height maturity. The correlation between predicted IQ's and measured IQ's was .83 with a standard error of .05, a figure which is certainly comparable to the correlations between IQ's obtained from different intelligence tests.

The result reported by Millard will be surprising to psychologists familiar with correlations between mental and physical characteristics. Jones (*4*, p. 601), commenting on a study dealing with physical-mental relationships by Honzik and Jones (*5*), wrote: "Correlations close to zero were found at each age from twenty-one months to seven years, but the uniformly positive coefficients for both boys and girls suggested the presence of a genuine, even if meager, relationship; there was, moreover, a slight tendency for growth rates in height or weight to be correlated with increments in mental scores." Both Olson and Millard are aware of this material, but their postulates imply a much higher degree of relationship (*8*, p. 169, and *13*, p. 31).

It must be admitted that there is little direct evidence on the particular relationship discussed by Millard; namely, that between height maturity and educational maturity. However, a partial explanation may be found in his paper: "Knowing only the mean IQ of the group, the problem became one of predicting individual IQ's from the height measures which covered a three-year period of measurement" (*7*, p. 446). How accurately could the IQ be predicted from 'height measures' if the investigator did not know the mean IQ of the group for whom the predictions were to be made? It appears that in order to make predictions of this degree of accuracy it may first be necessary to obtain the IQ's of the subjects from a general intelligence test, unless, of course, the investigator knows that he is dealing with some specific sample, such as, an average group. Further studies involving larger samplings should be used to investigate this problem. Probably the really important point in this connection is the inference that all growth patterns are closely interrelated.

PHYSICAL MATURITY AND READING READINESS

Millard points out that present attempts to use a child's growth status to estimate reading readiness are unsatisfactory because of the nature of the measures employed to determine maturity. Educational practices could be placed more in the realm of an exact science, he believes, if certain types of relationships were to be discovered between physical and educational 'growth' (*8*, pp. 43-44). For instance, it would be possible to use height to estimate the time to initiate reading instruction if there were a suitable mathematical relationship between 'growth' in height and reading. According to Millard, evidence for such a relationship is available.

From graphs showing height and progress in reading he states: "For example, in the three cases, each child acquires a height maturity of sixty-five to seventy-two per cent before reading begins" (*8*, p. 42); hence instruction in reading should not be initiated until the child has reached about two-thirds of his mature height. The meaning and significance of this observation will be considered.

1) Millard presents three sets of graphs for each of the three cases. In Figure 1 (a), CA was plotted against reading age; in (b), the data on reading age were smoothed before being plotted; in (c), CA was plotted against "percentage of development". Note that in (a), the lowest age at which the reading age was plotted was about ninty-five months. In the process of 'smoothing' the lowest age became eighty-two months. Finally, in (c), the youngest age shown on the graph is sixty-seven months. Apparently this change from ninty-five months to sixty-seven months was produced by extra polating beyond the data obtained by measurement.

It is legitimate to ask how much reliance should be given to such an estimate of the age of 'beginning reading'. In this particular instance, sixty-seven months may be a reasonable estimate of zero ability in reading as measured by a reading test, but one suspects that this child with an IQ of 173 might well have started reading before he was sixty-seven months old, for at that time his mental age would be in the neighborhood of 116 months, almost ten years. In another illustration presented by Millard, (*8*, p. 49), the estimated age of beginning reading was about 82 months (almost seven years) for a child with an IQ of 114. We might ask about the meaning of the expression 'age of beginning reading'. Surely, in view of the complexity of the reading process, it is difficult to define this age. Is it when the child first recognizes letters as naming pictures, or is it when he reads pictures (*13*, p. 121), or is it at some still earlier point in chronological time?

If there is some error in estimating the age of 'beginning reading', then, too, there will be an error in the conclusion that "each child acquires a height maturity of sixty-five to seventy-two per cent before reading begins."

2) The computation of 'percentage of development' for height (*8*, p. 47) employs some measure of maturity in height, and of course the percentage at any given age will depend upon the base used. The maximum age for which data were shown graphically by Millard was about fourteen years. Was the 'percentage of development' in height computed by using height at age fourteen as the base? If so, it would appear that the generalization should be to the effect that the child had attained about two-thirds of his fourteen-year-old height before he began to read. This is somewhat different from the reported conclusion, for typically children have not finished growing in height, and so have not reached 'height

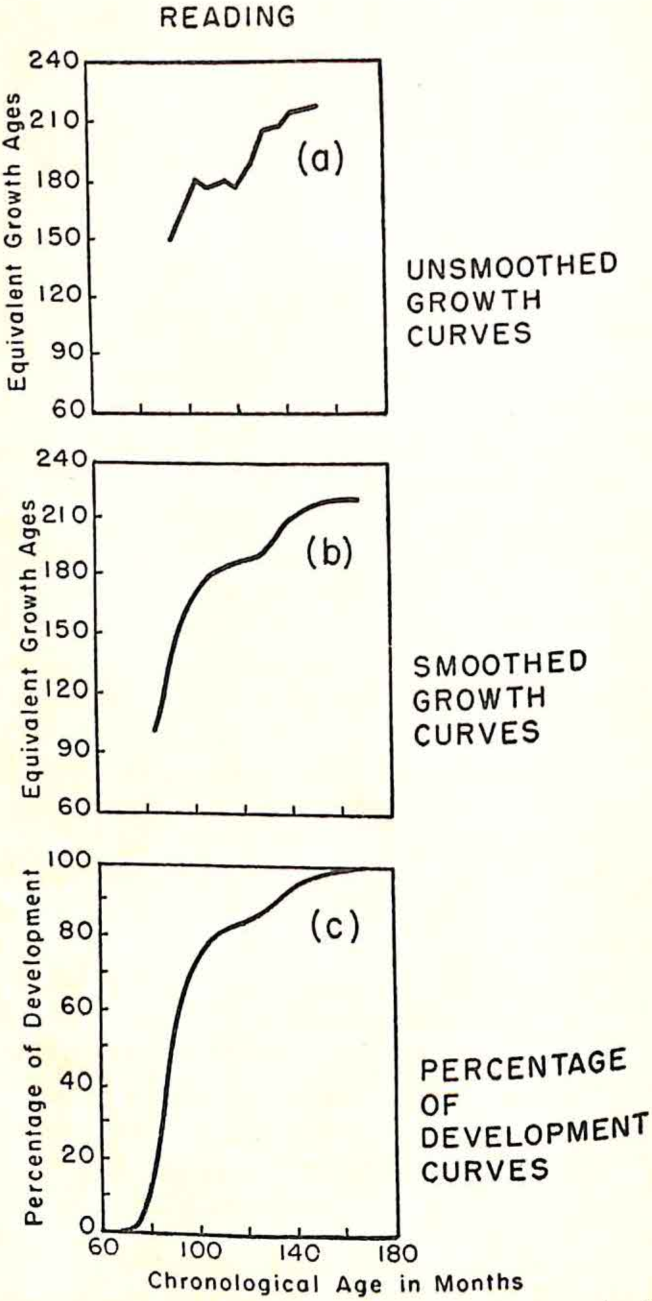


FIG.1. Three types of curves for achievement in reading for one child—Case 63F in Millard (*8*).

maturity' at age fourteen. Moreover, at age fourteen they (especially boys) are quite heterogeneous with respect to their relative maturity.

3) Even if such a relationship were found to hold for all children in the sample from which the three cases were selected, it does not necessarily follow that reading should start when the child has attained approximately two thirds of his 'height maturity', and not until then. Such an implication would fail to take into account the fact that these children were in a specific cultural, and especially educational, environment. Possibly the relationship claimed by Millard is simply a reflection of the existing educational program. In that case the data do not provide evidence that reading instruction should be initiated when every child has reached some specific fraction of his 'height maturity'.

4) Just how useful is Millard's index of height maturity for predicting reading readiness? Such an index is not readily available to the classroom teacher, for how can he know when a child has attained about two thirds of his height maturity? With what degree of accuracy is it possible to predict percentage of 'height maturity' in the absence of knowledge of mature height? Bayley (*2*) has recently provided a more accurate technique for making such predictions, but the method requires the use of information obtained from x-ray photographs. Such data are not usually available to educational authorities.

Apart from any of the above criticisms of this concept, the writer is inclined to believe that Millard's statement would have been better phrased in the past tense: "For example, in the three cases, each child had acquired a height maturity of sixty-five to seventy-two per cent before he began to read". This might have been a statement of fact; but it does not necessarily imply that reading should be postponed until a child has attained a certain degree of 'height maturity', on the hypothesis that he will not begin to read prior to that time. Further evidence is required for such generalizations.

THE CYCLICAL NATURE OF GROWTH

Millard assumes that 'growth' curves of both physical and educational characteristics are cyclical in nature, and claims that "a child who enters puberty late should show a late adolescent spurt in all kinds of growth. A child who is precocious physically should show early starting points in all of his learning curves" (*8*, p. 45). The educational implications of this principle were not as clearly enunciated by Millard as they were for some of the other principles under discussion. However, by inference at least, such a generalization should make it possible to predict the onset of spurts in educational progress from a knowledge of growth in physical characteristics. In this case, educational programs might well be synchronized with physical growth; the rapid growth in height at adolescence would then mean that we might also expect sharp increases in the learning of the materials of formal education. Somehow educators should be able to capitalize on such a relationship—if it really is a fact. The significance of this principle is, of course, clear for questions concerned with the interrelationships of 'growth' processes.

What is the evidence for the cyclical nature of educational 'growth'? Millard presents several growth curves for each of three children with CA plotted against 'percentage of development'. On the basis of these curves Millard postulated the principle of the cyclical nature of various kinds of 'growth'. He concluded that there were two cycles of growth between ages five and fourteen, the ages shown on his graphs. In his discussion, Millard refers to the "starting point" of "preadolescent learnings" and of "all adolescent growth." He estimates the average ages at which each of the three cases "started their preadolescent learning," and their "adolescent growth," together with a measure of "average deviation of starting points."

1) For one of the three pupils, zero percentage of development in social science was about sixty months, whereas from a previous figure showing his 'growth' it appeared that the earliest age for which any data on 'growth' in social science were available was about ninety-seven months (*8*, Figures 5 and 12, pp. 37 and 47). Apparently the beginning of 'preadolescent learning' was obtained by extra polation. The validity of this estimate is surely open to question when it is obtained by such means. The same criticism applies to the use of an estimated age of maturity, since it was used in computing the 'percentage of development' and since it was apparently obtained by extra polation.

2) The present writer finds some difficulty in determining by inspection the location of the 'starting points of adolescent growth' in the graphs presented by Millard. The presence of a spurt is in doubt in nine of fifteen curves showing 'growth' in educational

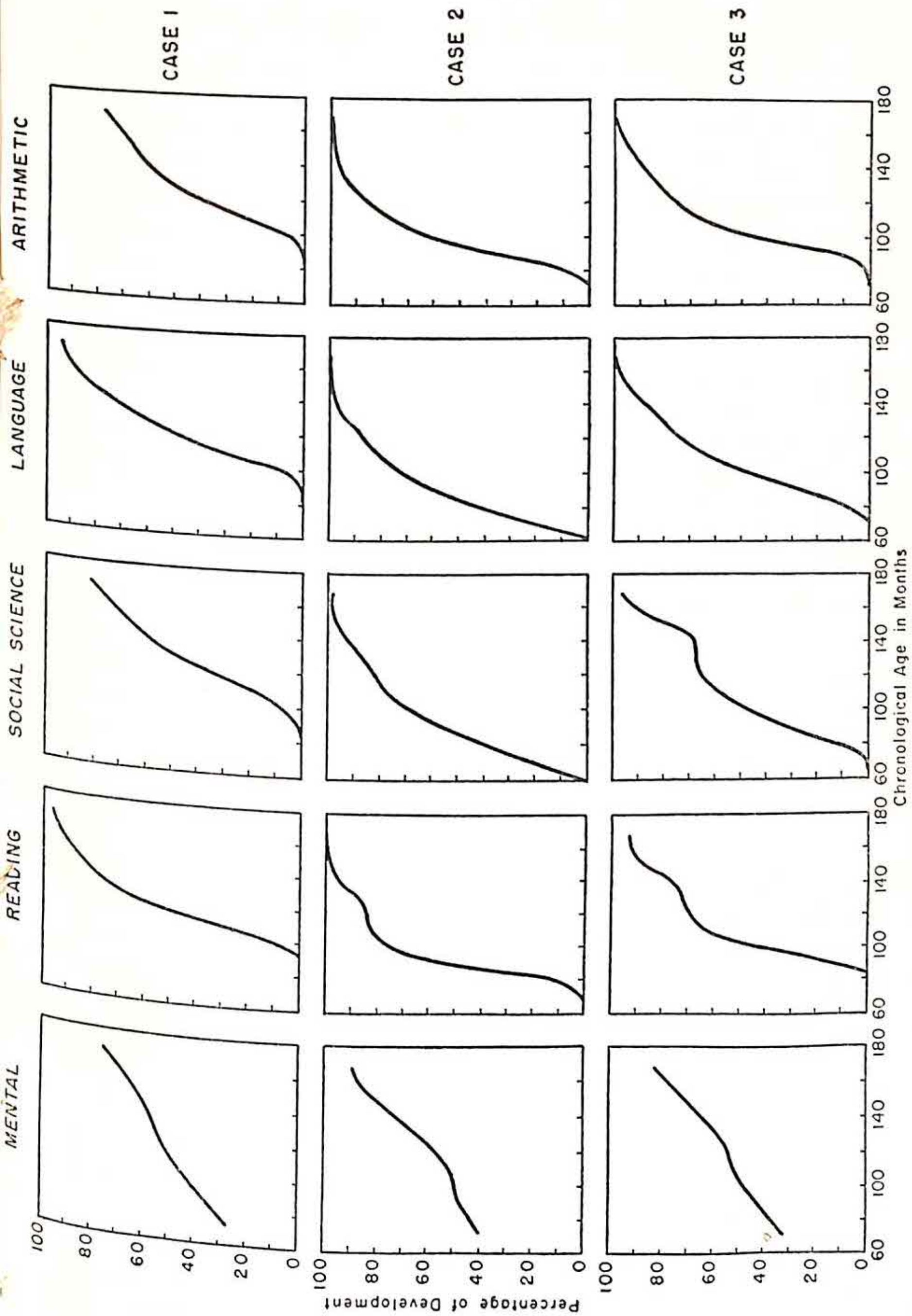


FIG. 2. "Cyclical growth" for three cases presented in Millard (*8*)

characteristics; and the minor 'jogs' in some of the other curves may have been produced by the unreliability of the measuring instruments. Of course, the original tabular material, larger graphs, or mathematical equations might reveal a spurt in these processes at adolescence, but the data as presented do not. Further, the age of onset of adolescence is not apparent from Millard's figures or his descriptive materials. These comments do not demonstrate that 'educational growth' is not cyclical in nature; but they do imply that the data presented by Millard are not satisfactory evidence of such a pattern of educational change.

THE PRINCIPLE OF FAMILIAL RESEMBLANCE

Evidence of familial resemblance is presented in graphical form by Olson the Hughes in Barker, Kounin and Wright. "The siblings available to our study show a high degree of similarity in the pattern of growth . . . . The illustration employed for brothers born thirty-three months apart, shows striking similarities in level and pattern for the six variables" (*1*, p. 205). Some of Olson's graphs are reproduced in Figure 3.

Similarity in the pattern of growth for some of the variables shown in the graphs is admitted if similarity refers to the general shape of the curves, but not if it implies that they are coincidental; this is true in some instances, but in others sibling differences of as much as twenty-four months occur. Little serious attention has been given to an analysis of patterns of growth; Olson does this by inspection. Some comparability in level of growth for siblings is to be expected, since siblings correlate about .50 in mental ability. A more systematic consideration of the problem might provide valuable information about the nature of growth processes.

The curves of Figure 4 have also been submitted as evidence that there is familial resemblance in educational 'growth'. Notice that the reading curve for one child (B″) is below that of his sibling (B) throughout the age range six to twelve years. Since the curves are roughly parallel, it might be argued that there is similarity in the pattern of 'growth' in reading. What of 'level'? Consider the reading ages of B and B″ at age twelve. The RA's are about 197 and 164 months respectively. Is this evidence for striking similarity of level of growth?

Surely serious application of this principle could lead to situations which Olson would deplore: "Conflicts between expectancies and the child's growth are a frequent source of social and emotional repercussions" (*11*, p. 5). Child B″, in Figure 4, is approximately thirty months, two and one-half years, behind B in reading at

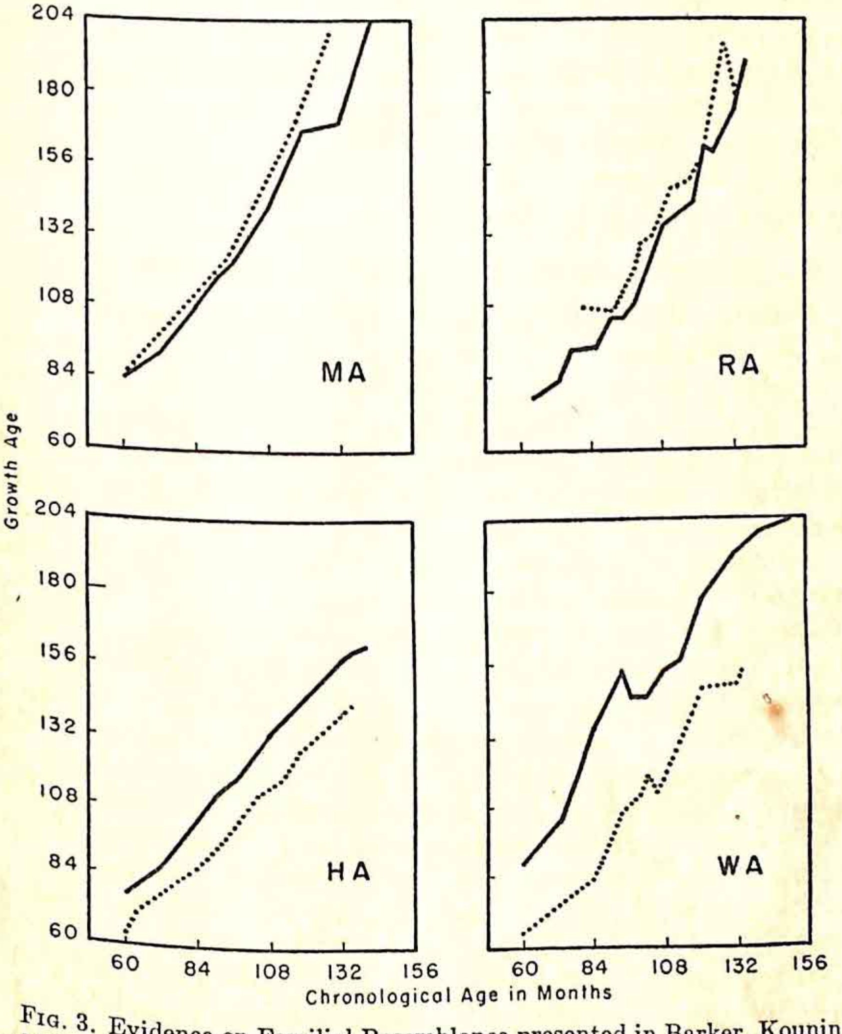


FIG. 3. Evidence on Familial Resemblance presented in Barker, Kounin and Wright (*1*).

twelve years of age, after approximately six years of formal education. If the parents had similar expectations for the two boys, as the principle of familial resemblance would suggest, then B″ might not have a very 'serene and productive' life either at home or at school.

DISPLACEMENT, CONVERGENCE, AND PERTURBATIONS

Olson suggested the hypothesis that "an environmental perturbation may be reflected with greater or less immediacy in the various structures and functions that are sampled in attempting to describe the total growth of the child" (*1*, p. 205). A part of the

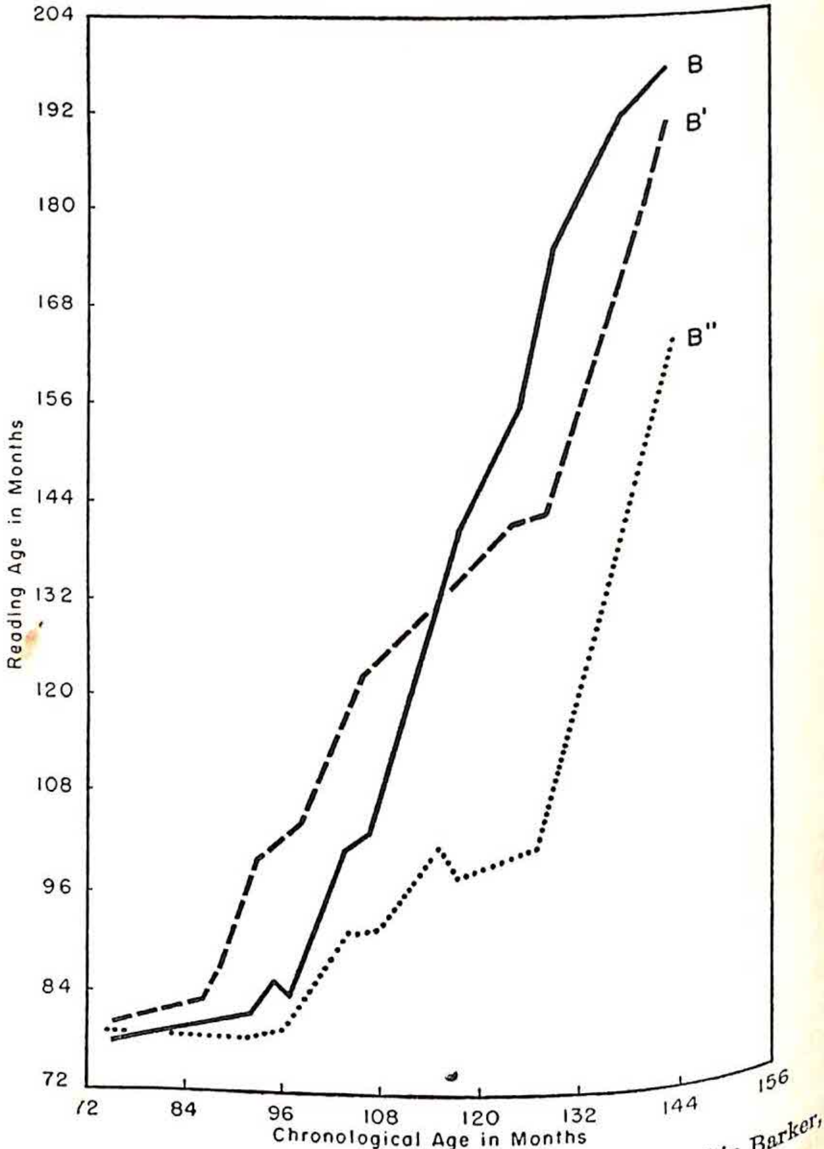


FIG. 4. Evidence on sibling resemblances in reading presented in Barker, Kounin and Wright (*1*).

graph used as evidence for this principle is reproduced in Fig. 5. It is recognized that Olson has used several qualifying expressions in the formulation of the principle relating to environmental perturbations: may, greater and less. Possibly the principle must be accepted as it is stated, but certain comments seem relevant.

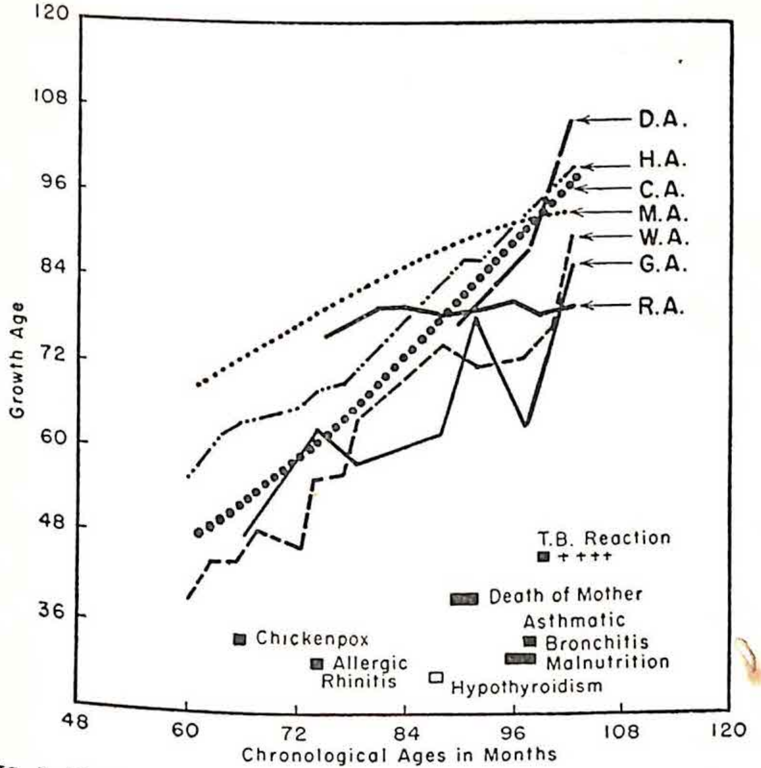


FIG. 5. Evidence of the effect of perturbations on growth presented in Barker, Kounin and Wright (*1*).

Olson drew attention to the change in the weight curve after about eighty-four months of age, suggesting that the change might be attributed to certain environmental factors operating at that time. However, it should be noted that there were no comparable changes in the curves depicting other types of growth, except possibly in the case of grip age. In this latter case, the irregularity after age eighty-four months might be more a matter of motivation than of any underlying change in growth in strength of grip. Possibly the 'hammering down' (to use Olson's phrase) by environmental shocks was more the exception than the rule in the case of this child. It is worth observing, also, that the growth curve for weight was irregular even prior to the onset of the specific environmental shocks to which Olson draws attention.

Honzik, Macfarlane and Allen (*6*) reported that it was difficult to find environmental factors which would account for marked changes in mental growth. In some cases environmental forces could be correlated with changes in rate of growth, but not in others; sometimes environmental deprivations occurred at the same time as improvements in the rate of mental growth. "Children whose mental test scores showed the most marked fluctuations had life histories which showed unusual variations with respect to disturbing and stabilizing factors. However, there were other children whose scores remained constant despite highly disturbing experiences" (*6* p. 315).

On the basis of the curve shown in Figure 5, Olson proposed the principles of the resistance to displacement and the tendency toward convergence. According to these principles, a child tends to resist any external attempts to modify his rate of growth (which, for Olson, includes behavior and achievement) and to resume his normal rate of growth after cessation of these attempts to change his growth pattern. What is meant by the term normal in this context? Apparently it refers to the growth which occurs in the absence of any factor employed to produce a change. Surely it is reasonable to expect a change in the rate of growth (including learning) when any specially-introduced factors no longer operate. This is the essence of teaching, for teaching is reasonably regarded as an external attempt to modify the educational achievements of children in our society.

In discussing 'normal growth' Olson seems to imply an innate guiding force when he states: "Unfolding design appears inherent in the nature of a species, and the level and rate of emergence is also dependent on immediate ancestry" (*13*, p. 40). This may be more or less true of many phylogenetic characteristics, but it may be doubted that it is a factor of primary importance in learning geographical and arithmetical facts. The term 'normal growth' must, in addition, take into account the nature of the environment, especially if growth is to include learning of the type that occurs in our schools. It is 'normal' for a certain sea organism to develop two eyes in a specific type of environment; but it is just as 'normal' for it to develop a single eye when the environment is modified in a certain direction (*18*, p. 7). 'Normal' growth means normal for a certain organism in a specific environment. From this point of view, teaching methods, including diagnostic and remedial procedures, readiness programs, etc., assume more, rather than less, significance in the educational 'growth' of the individual.

A study by Tilton has been cited as evidence on the inefficacy of remedial procedures for producing changes in the rate of educational 'growth' (*17*). It is questionable whether he would agree with some of the interpretations of his results. Some of the pupils showed significant gains as a result of special effort on the part of the teacher. The place of arithmetic in an educational profile was improved significantly (statistically and educationally) for ten of thirty-seven pupils; there were losses for some of them. The failure to find a larger number of significant gains might be attributed to the fact that the teachers did not plan and execute an effective remedial program. Tilton apparently suspected such a condition, for he suggested that it is "conceivable that poor remedial effort may do more harm than good." It is unreasonable to accept Millard's conclusion that "Tilton's study shows the futility of diagnostic and remedial activities in producing permanently improved achievement" (*8*, p. 51). Certainly there were changes in the educational profiles of some of the children in some of the classrooms. The average gain may have been approximately zero—but this could arise from gains being balanced by losses. Certainly the term 'permanent' cannot be applied to the results of Tilton's study, which required a rather short period of time and which involved no long-time follow up.

SUMMARY

Some of the facts, concepts, and principles of growth and development of children found in several papers and two recent textbooks were selected for consideration and analysis. These concepts were analyzed from several points of view:

a) statistical bases and meaning;
b) psychological meaning;
c) validity of the interpretation of the basic data;
d) educational implications.

It was pointed out that the greater predictability of organismic age compared with other types of ages was a statistical artifact; this age, being the mean of a series of other ages, can be expected to have greater stability. The statistical meaning of the OA was questioned on the grounds that there is no evidence that age units are equal. The significance of certain types of ratio comparisons is in doubt since the location of absolute zero on the scale has not been established. The psychological meaning of certain concepts was questioned, as, for instance, the meaning of the average of a series of age scores of such diverse structures and functions as height, dentition and skill in arithmetic.

In certain cases, it seemed doubtful that proposed conclusions necessarily followed from the data. For example, there is some question about the presence of cycles in educational 'growth'. Finally, the validity of some of the proposed educational implications was questioned, particularly with reference to instructional methods and readiness programs. Of course, neither Olson nor Millard really means that teaching methods and materials are not important: "The second implication relates to the tremendous responsibility of the school for providing the enrichment so essential for maximum language development. From this point of view, the schools that use explorations, broad area projects, new experiences of all kinds, new books, new observations are providing an educational program in harmony with a great majority of the research findings" (*8*, p. 171). The fact that educational planning must consider genetic and maturational as well as environmental factors in growth cannot be refuted. A one-year-old child cannot read or compute or play tennis or chess or the piano. On the other hand the ability to do these things is not simply a matter of allowing a design to unfold during the passage of time.

CONCLUSIONS

The basis and the meaningfulness of certain concepts have been questioned; this does not mean that many of the proposals for educational materials and methods are erroneous and fallacious. To raise questions, however, is not to imply that educators can safely ignore the facts and generalizations of growth and development. Buswell recently drew the attention of educators to the importance of an understanding of the learner for curriculum planning. However, he was impelled to observe that much of the work in the field of growth and development, and that many of the suggestions for education were "characterized by sentimentalism and lack of valid techniques. But the movement embodies the kernel of a significant idea and has important implications for the curriculum worker" (*9*, p. 462).

The principles proposed by Olson and Millard may be sound; but the possibility of invalidity in the evidence and in the interpretations has been suggested. Empirical studies might reveal that such concepts as the age unit and the organismic age have the values claimed for them despite their logical weaknesses. Analyses of longitudinal data to test specific hypotheses arising from the questions raised in this paper are necessary. Investigations are urgently needed in view of the educational importance of the facts, principles and concepts about the growth and development of children. Various hypotheses using longitudinal data are now under investigation at the Institute of Child Welfare, University of California. Such studies should provide further evidence on the nature and interrelationships of physical, mental, educational, social and emotional changes that occur in the first eighteen years of life.

REFERENCES

1) R. G. Barker, J. S. Kounin, and H. F. Wright. (Eds.). *Child Behavior and Development*. New York: McGraw-Hill, 1943.

2) Nancy Bayley. "Tables for predicting adult height from skeletal age and present height," *The Journal of Pediatrics*, 28: 49–64, 1946.

3) ——. "Consistency and variability in the growth of intelligence from birth to eighteen years," *The Journal of Genetic Psychology*, 75: 165–196, 1949.

4) Leonard Carmichael (Ed.). *Manual of Child Psychology*. New York: John Wiley & Sons, 1946.

5) M. P. Honzik and H. E. Jones. "Mental-physical relationships during the preschool period," *Journal of Experimental Education*, 6: 139–146, 19[illegible]7.

6) M. P. Honzik, J. W. Macfarlane, and L. Allen. "The stability of mental test performance between two and eighteen years," *Journal of Experimental Education*, 17: 309–324, 1948.

7) Cecil V. Millard. "Further comments on the November issue," (IQ's) *Educational Method*, 19: 445–447, 1940.

8) ——. *Child Growth and Development in the Elementary School Years*. Boston: D. C. Health and Company, 1951.

9) National Society for the Study of Education, Forty-first Yearbook, Part II, *The Psychology of Learning*. Bloomington, Ill.: Public School Publishing Company, 1942.

10) Quinn McNemar. *The Revision of the Stanford-Binet Scale*. Boston: Houghton-Mifflin, 1942.

11) Willard C. Olson and Bryon O. Hughes. *Manual for the Description of Growth in Age Units*. Ann Arbor, Michigan: The Edwards Letter Shop, 711 N. University, Ann Arbor.

12) ——. "The concept of organismic age," *Journal of Educational Research*, 35: 525–527, 1942.

13) Willard C. Olson. *Child Development*. Boston: D. C. Heath and Co., 1949.

14) M. W. Richardson. "The logic of age scales," *Educational and Psychological Measurement*, 1: 25–34, 1941.

15) Frank K. Shuttleworth. *The Physical and Mental Growth of Girls and Boys Age Six to Nineteen in Relation to Age at Maximum Growth*. Washington, D. C.: Society for Research in Child Development, 1939.

16) L. M. Terman and Maud A. Merrill. *Measuring Intelligence—A Guide to the Administration of the New Revised Stanford-Binet Tests of Intelligence*. Boston: Houghton-Mifflin, 1937.

17) J. W. Tilton. "An experimental effort to change the achievement test profile," *Journal of Experimental Education*, 15: 318–323, 1947.

18) Kimball Young. *An Introductory Sociology*. New York: American Book Company, 1934.

THE CONTRIBUTIONS OF FILM INTRODUCTIONS AND FILM SUMMARIES TO LEARNING FROM INSTRUCTIONAL FILMS[1]

C. W. LATHROP, JR., C. A. NORFORD AND L. P. GREENHILL

The Pennsylvania State College

This exploratory study is concerned with investigating the contributions to learning of some introductory and summarizing sequences in existing instructional films. The study is divided into two parts: Part I deals with film introductions; Part II deals with film summaries.

I. FILM INTRODUCTIONS

A 'Film Introduction' is defined as that portion of a film, excluding the main and credit titles, which begins the presentation, and runs up to the beginning of the body of the film. The purpose of this part of the study was to measure the effect on learning of some introductory sequences in existing instructional films.

One hundred and thirty instructional films containing introductory sequences were viewed and classified according to (1) the functions which introductions appeared to perform, and (2) the film techniques employed. Eleven different functions of a film introduction were identified. They are listed in order of frequency of use.

1) Stresses the importance of the material in the film.
2) Poses the problem to be dealt with in the film.
3) Introduces the characters to appear in the film.
4) Sets the stage, that is, orients the audience to the scene of the action.
5) Gets attention of the audience by some dramatic device.
6) Points out important features which will be developed in the film and to which the audience should pay special attention.
7) Stresses the consequences if the material in the film is not learned.

---

[1] This research was conducted by the Instructional Film Research Program, The Pennsylvania State College, under Contract N6-onr-269, Task Order VII, with the Special Devices Center of the Office of Naval Research.

8) Shows the learner the relevance of the material in the film to what he has learned previously.
9) Explains to the instructor the situation for which the film is intended.
10) Provides additional inspiration which might motivate the student or trainee to undertake further activities after seeing the film.
11) Shows the purpose of the film.

Twelve different film techniques were used in the presentation of the introductory sequences in the one hundred and thirty films reviewed. They are listed in order of frequency of their use.

1) Live action
2) Narration by an off-stage commentator
3) Demonstration of a task being performed
4) Diagrams, still shots, tables, graphs, maps
5) Animation
6) Use of models (scale representations)
7) Titles to explain the film, etc.
8) Flashes forward (short shots of scenes to follow are included in the introduction).
9) Remarks by an authority on the subject
10) Slow motion or speeded motion
11) 'Dramatic' live action (action used with dramatic effect).
12) Audience participation (as in asking a question and allowing time for an answer).

THE EXPERIMENTAL PROCEDURE

The Films.—From the films reviewed three were finally chosen by a panel of instructors as having what seemed to be the best introductory sequences. Their characteristics are given in Table I.

The Experimental Versions.—Two experimental versions were prepared for each of the three films: Version I was the complete film; Version II was the same film minus the introductory sequence only. The preparation of the 'no introduction' versions was a comparatively easy matter as, in each film, there was a fade-out in the visuals and a break in the sound track between the credit titles and the introduction, and between the introduction and the main body of the film. The main title and credit titles were included in both versions of each film.

The Tests.—Tests were constructed on the material in each of

TABLE I. CHARACTERISTICS OF FILMS USED IN THE STUDY OF FILM INTRODUCTIONS

Title	Total Film Length	Length of Introduction	Functions of Introduction	Film Technique used in Introduction	Subject Type
Sulphur and Its Compounds	387 ft. 10 min. 45 sec.	65 sec. 10.1% of total	Stressing importance Introducing characters Setting the stage	Live action Models Title Narration	General Science-Chemistry
Mammals of the Rock Mountains	369 ft. 10 min. 15 sec.	45 sec. 7.3% of total	Setting the stage	Live action Animation Narration	General Science-Biology
Rivers of the Pacific Slope	414 ft. 11 min. 30 sec.	25 sec. 3.6% of total	Stressing importance Setting the stage Posing the problem	Live action Narration	General Science-Geography

the three films, bearing on the facts in the body of the film only; no questions were asked on the facts contained only in the introduction. Multiple-choice test questions each with four choices were asked on three different classes of facts in the films:

1) the facts contained in the visuals only
2) the facts contained in the sound track only
3) the facts found in both the visuals and sound track

Pilot-runs were made to check the reliability of the tests, and poor questions were eliminated. Each test finally included from fifty to sixty items.[2]

Test Population.—Approximately five hundred ninth-grade high-school students from two schools took part in the experiment. Good coöperation from the schools made it possible to achieve a practical degree of randomization of the entire ninth-grade population in each school into three groups.

One group acted as a control group and took the test without seeing the film. The second group saw the complete film (Version I), while the third group saw the 'no-introduction' version (Version II). The groups were rotated so that each group became a different experimental group for each of the three films. Thus, each group acted as a control group for one film, as a group seeing the version without the introduction for another, and, finally, as a group seeing the entire film for the third. The groups were also rotated with respect to projection rooms and test administrators.

The test followed immediately upon the film-showing.

RESULTS OF THE EXPERIMENT

The results of the experiment are summarized in Table II. These results show that the groups which saw the experimental films generally did somewhat better than the control groups which did not see the films. However, the differences between the groups which saw the entire film, and those which saw the film minus the introduction were small. For two films the introductions apparently made small positive contributions ('Sulphur and Its Compounds' +1.14, 'Rivers of the Pacific Slope' +1.81), while for the third film, 'Mammals of the Rocky Mountains', the introduction apparently had an adverse effect on learning, the difference between experimental groups being −2.55. This latter unexpected result was carefully checked and proved to be authentic.

---

[2] Kuder-Richardson coefficients of test reliability were as follows: Sulphur and Its Compounds .57; Mammals of the Rocky Mountains .69; Rivers of the Pacific Slope .72. These are minimum estimates of reliability.

The results indicate that among existing films, typical introductory sequences can make small positive contributions to learning, while in other instances introductions may have an adverse effect on learning, possibly through misdirecting the student's attention.

TABLE II. SUMMARY OF TEST SCORES

	Control Group	Film minus Introduction Group	Complete Film Group	Difference Attributable to Introduction
(a) "Sulphur and Its Compounds"				
Number of Subjects	168	166	168	+1.14*
Mean Score	16.97	21.61	22.75	
Standard Deviation	3.43	5.45	5.36	
(b) "Mammals of the Rocky Mountains"				
Number of Subjects	168	171	174	−2.55***
Mean Score	22.55	31.23	28.68	
Standard Deviation	4.93	7.27	6.63	
(c) "Rivers of the Pacific Slope"				
Number of Subjects	165	167	164	+1.81**
Mean Score	16.23	22.96	24.77	
Standard Deviation	4.10	5.95	6.70	

* Significant at the 6% level of confidence
** Significant at the 1% level of confidence
*** Significant at the 0.2% level of confidence

II. FILM SUMMARIES

The term 'Film Summary' as used here, refers to a concluding sequence produced as an integral and purposeful part of an educational film. It embraces one or more of the functions of review, recapitulation, statement of importance, and the issuing of a challenging note; it may also contain an 'application' of the information given in the film. The film summary is usually preceded by a fade in the visuals, and a natural pause in the sound track, which separates it from the body of the film. It does not include'The End' title.

The purpose of this part of the study, which paralleled the study of film introductions, was to measure the effect on factual learning of the summary sequences in some existing instructional films.

Eighty-seven instructional films, which included summaries, were analyzed and classified according to the purpose for which the summary appeared to be used, the film techniques employed, and the length of the summary in relation to the total length of the film.

Six different functions served by film summary sequences were identified. They are listed in order of frequency of use.

1) *Review.*—A mere topical outline of the film; a brief restatement of the organization of the film rather than of its informational content.

2) *Application.*—The illustration of a point of information by a concrete example.

3) *Importance.*—Stressing the value to the viewer of the information in the film.

4) *Challenging note.*—The issue of a challenge to the viewer to apply the information in the film, or to seek further information, or to undertake other activities. This heading would also include the posing of questions for thought or discussion.

5) *Recapitulation.*—A brief repetition or restatement of the principal points in the film.

6) *New information.*—The summary may contain information not previously given in the film, or it may relate the film to new material to follow.

The film techniques used in the presentation of the eighty-seven film summaries were also listed. They are given here in order of frequency of use.

1) *Live action.*—Simple movement as from real life.

2) *Narration.*—The off-stage voice of a narrator.

3) *Flash backs.*—The reshowing of parts of scenes used in the body of the film.

4) *New scenes.*—Scenes not shown previously in the film.

5) *Music.*—Musical background behind commentary.

6) *Animation.*—Use of drawings and charts, etc. with movement.

7) *Questions.*—Asking question, either by titles, or narration.

8) *Lip-synch.*—A person on the screen speaking, with synchronous recording of the speech.

9) *Still shots.*—Photographs or drawings or diagrams without movement.

THE EXPERIMENTAL PROCEDURE

The Films.—Three films were finally selected from the eighty-seven reviewed, as having what appeared to be the best summary sequences. The characteristics of these films are given in Table III.

The Experimental Versions.—For each of the three films, two experimental versions were prepared: Version I was the complete film, Version II was the same film minus the summary sequence. The 'End' title was retained in each version.

The Tests.—Tests were constructed which were based on the information in the body of the film, and not on items appearing only in the summary. Multiple-choice questions with four alternatives were used, together with a proportion of true-false questions. A pilot study was made to determine the reliability of the three tests, and the tests were revised and proved for use in the final study.

The test on *The Cell* contained fifty-eight questions, eleven of which were true-false; the test on *Magnetism* contained sixty questions, eight of which were true-false; and the test on *Rivers of the Pacific Slope* contained fifty-two questions, all of which were of the multiple-choice type.[3]

Test Population.—Five hundred and sixty-one ninth-grade students from three Pennsylvania high schools were tested in this experiment. Good coöperation by the schools made it possible to achieve a practical degree of randomizing by splitting the entire ninth-grade population of each school into three experimental groups.

As in the study of Film Introductions, one group acted as a control group and took the test without seeing a film, while a second group saw the complete film (Version I), and the third group saw the film minus the summary (Version II).

The groups were rotated so that each group became a different experimental group for each of the three films. The rooms for film showings and the test administrators were also rotated to distribute any differences which may have arisen from these variables.

The test followed immediately on each film showing.

---

[3] Kuder-Richardson coefficients of test reliability were as follows: The Cell .82; Magnetism .83; Rivers of the Pacific Slope .69. These are minimum estimates of reliability.

RESULTS OF THE EXPERIMENT

The results of the experiment are summarized in Table IV. These results indicate that the groups which saw the films did definitely better on the tests than the control groups which did not see the films. The differences were small between the groups which saw the complete films, and those which saw the films minus the summaries. For all three films the summaries made small positive contributions to learning, the differences in favor of the films with summaries being as follows:

TABLE IV. SUMMARY OF TEST SCORES

	Control Group	Film minus Summary Group	Complete Film Group	Difference Attributable to Summary
(a) "The Cell: Structural Unit of Life"				
Number of Subjects	192	184	185	
Mean Score	24.67	33.00	33.57	+0.57
Standard Deviation	5.65	7.19	8.52	
(b) "Magnetism"				
Number of Subjects	184	185	192	
Mean Score	32.94	37.00	38.93	+1.93*
Standard Deviation	8.99	8.66	8.57	
(c) "Rivers of the Pacific Slope"				
Number of Subjects	185	192	184	
Mean Score	17.10	24.95	25.25	+0[illegible]
Standard Deviation	4.34	6.86	6.30	

* Significant at the three per cent level of confidence.

The Cell: Structural Unit of Life	+ .57
Magnetism	+1.93
Rivers of the Pacific Slope	+ .30

It should be noted that only one of these differences (for the film *Magnetism*) reached accepted levels of statistical significance. The results suggest that these films, which include what seemed to be the best available summary sequences as produced today, are not materially better than they would be without the summaries.

DISCUSSION AND CONCLUSIONS

Failure of the introductions and summaries to contribute significantly to learning from the films used might be attributed to one or more of the following factors:

1) *The short time devoted to introductory and summary sequences.*—The average length of time devoted to summaries or introductions was 9.6 per cent of the total film running time. The shortest introductory or summary sequence was 3.6 per cent of the film-running time; the longest was 13 per cent.

2) *Learning from the films themselves (without summaries or introductions) was comparatively low.*—The highest mean learning increment was only nine points on a 58-item test. This low level of learning could be the result of the fact that all of the films used in this experiment presented a considerable amount of factual information in a very brief time. (A fifty- or sixty-item test was readily constructed on each ten-minute film.) Under such conditions, it seems possible that no short summary or introduction could contribute significantly to learning from the films. This is especially true, perhaps, when we consider that summaries and introductions are not mere positional designations; they refer to elements of the film which function in integration with the main body of the film. It is likely that they do not make a simple addition to the film's effectiveness. Their effectiveness is no doubt conditioned by the characteristics of the main film. The rapid rate of development in the films studied, then, may very well have operated against the effectiveness of summaries and introductions. It might be hypothesized that the 'better' the film, the more effective the summary or introduction can be.

3) Although the films studied in this experiment were selected because their introductory and summary sequences appeared to incorporate useful teaching principles, there is no assurance that they represented the most effective application of those principles; nor is there any assurance that the summaries and introductions studied did, in fact, represent the 'best' summaries of their lengths and kinds.

4) Where the function of a summary is to increase discrete factual learning, the effectiveness of the summary is obviously limited by the number of pictures and the amount of commentary which it contains, and the time that is devoted to it. In such a case, the effectiveness of the summary, measured by a test of discrete

factual knowledge, would seem to be necessarily limited. Where the summary is used, however, to restate principles, or to formalize and crystallize concepts that have been demonstrated or shown in the film, conceivably its effect on learning may be more significant.

5) With regard to introductions, it might be hypothecated that an introduction can be made to contribute significantly to learning from a film if it provides a motivational stimulus or 'set', or if it relates what is to be taught in the film to the previous experience of the audience. It is interesting to note that in very few of the films examined was the introduction used to state the objectives of the film.

In the light of this discussion, then, the findings of this experiment cannot be generalized; it cannot be concluded that introductions and summaries, in general, make no significant contribution to learning from films.

The experiment suggests that introductions and summaries cannot be considered as single variables; rather, they involve a complex of variables similar to those which determine the effectiveness of the film itself.

The length of introductions and summaries, the relationship they bear to the main body of the film, the degree to which they are integrated with the main body of the film, and the functions which they can perform are all important variables which should be systematically investigated in future studies.

GROUP THERAPY WITH RETARDED READERS

BERNARD FISHER

The Children's Village, Dobbs Ferry, New York

The importance of the problem of the retarded reader is indicated by the large time allotment that is assigned to remedial reading in the elementary school, the wealth of teaching devices that have originated in relation to it, and the numerous investigations that have been made in this field. Authorities are in agreement that most other school subjects are affected by reading ability. "These are admitted facts and are examples of an increasing number of findings that emphasize the value of establishing good reading habits." (*11*)

Failures in reading are frequent and many pupils never acquire adequate reading skills. Gates states: "Despite the quantity of experimental data, the wealth of ingenious teaching devices, the range of interesting children's reading material and the large amount of school time available for teaching reading, a surprisingly large number of pupils still experience difficulty in acquiring satisfactory reading skills." (*11*)

In one study it was found that reading was, "the most frequent cause of school failure." (*15*) Schonell (*18*) found that the amount of reading failure in ten different primary classes averaged 22.6 per cent. Betts (*4*) estimated that approximately ten per cent of American school children require special assistance because of reading disabilities. These estimates of the number of children troubled by reading failures indicates the magnitude of the problem.

The varied approaches recommended for the remediation of reading disabilities are usually related to the authorities' beliefs concerning the cause of the disability. These causes may be grouped under the general categories of physical, psychological, educational, emotional and social, or a combination of causes. Gann (*9*) reviewed a large number of representative investigations and points of view that stress particular functions or group of functions that produce reading difficulty. She concludes: "though the review presents points of view bearing a positive relationship to reading difficulty, numerous studies can be located in the literature negating these findings." She concludes, "the conclusiveness of the various approaches already existing cannot be established."

Despite this variety of inconclusive evidence, remedial reading teachers who use different approaches based upon the findings of the authority or authorities with whom they are acquainted usually achieve some degree of success. It has recently been suggested (*1*, *2*, *5*, *8*) that in addition to the improvement of reading techniques for the correction of reading disabilities, the psychotherapeutic relationship of the teacher, tutor, or clinician was a major factor in the correction of the reading disability. This suggests the possibility that all of the remedial methods may have had in common a psychotherapeutic relationship that arises when one person attempts to help another.

In support of this hypothesis, Axline briefly noted the varied attacks that had been made upon the reading problem. "(The varied attacks) gave the same results: successful in many cases, useless in others. These varied methods have one characteristic in common: the motivation of the teacher." (*2*) She described an experience with thirty-seven second-graders who were poor readers or non-readers and whose teacher tried to help them by using a non-directive therapeutic approach. She employed art materials, free dramatics, puppet plays, music, creative writing, etc.

Axline concluded by stating: "This study indicates that a non-directive approach might be helpful in solving certain 'reading problems.' It indicates that it would be worth while to set up research projects to test this hypothesis further: that non-directive therapeutic procedures applied to children with reading problems are effective not only in bringing about a better personal adjustment but also in building up a readiness to read." (*2*)

Bills described a study in which he worked with a class of third-graders who had previously been classified as slow learners. He concluded that: "1) Significant changes in reading ability occurred as a result of the play therapy experience, 2) personal changes may occur in non-directive play therapy in as little as six individual and three group play therapy sessions, and 3) there appears to be no common personality maladjustment present in (his) group of retarded readers." (*5*)

An earlier investigation into the relationship between the improvement in emotional adjustment and the improvement in reading, is Redmount's (*16*) 'before' and 'after' study of a resident six weeks' summer clinic for twenty-three children with reading difficulties. Reading tests showed that forty-eight per cent improved and twelve per cent seemed to regress. Rorschach tests indicated that thirty-nine per cent showed personality improvement, with twenty-six per cent adversely affected.

Many articles have called attention to the relationship between personality and the reading process. Strachey (*19*) from the psychoanalytic point of view, stated that reading represents a sublimation of oral tendencies, especially those of a sadistic and destructive nature. He feels that skill in reading breaks down when these oral drives are unstable or completely repressed and sublimated. Klein (*14*) suggested that for those children whose neurotic conflicts are largely concerned with trying to keep aggressive drives repressed, sublimated expression of reading may not be permissible to ego and super ego. A recent investigation by Vorhaus (*21*) tends to support these early hypotheses. Blanchard (*6*) described two groups of individuals who display reading disabilities: (1) the non-neurotic who develops emotional conflicts largely out of inability to read, (2) the neurotic whose emotional conflicts and difficulties in personality have preceded the reading difficulties which have arisen from these early maladjustments. Hardwick (*13*) stated that many reading cases show evidence of timidity and inferiority. Bennett (*3*) noted lack of persistence and attention, preference for solitary and inactive life, fears, indecisions, loneliness and crying spells among the retarded readers. Gates (*10*) pointed out that children with reading difficulties lack persistence, do not concentrate well, show marked sensitivity, are withdrawn, daydream and show a lack of aggressiveness, necessary for effective adaptation in learning to read. Tulchin (*20*) came to the conclusion that undesirable behavior patterns or personality maladjustment may be traced to reading disability. Witty and Kopel (*22*) found that fully fifty per cent of seriously retarded readers are characterized by fears and anxieties which require a program of reëducation aimed at the reëstablishment of self-confidence and the removal of anxieties.

Evidence like the above caused this investigator to hypothesize that if reading disabilities are in large part caused and effected by emotional adjustment, then psychotherapeutic steps would help to remove the disability. To test this hypothesis the following study was conducted.

DESIGN

The subjects were twelve residents of the Children's Village, an institution for delinquent boys. These twelve were part of a larger group which was receiving special remedial reading instruction. The subjects were all designated as having a reading disability on the basis of their past school history, case record and psychological test results. As is indicated in Table I they ranged in age from ten years, five months to twelve years, nine months. Their initial reading abilities, as designated by reading age, ranged from six years, seven months to eight years, ten months. They were all more than three years retarded in reading, had the same regular classroom teacher, and received remedial reading instruction for three hours each week from the same remedial teacher.

Table I. Test Data for Experimental Groups

	IQ	Age	Initial Reading Age	Final Reading Age	Gain in Months
Therapy Group					
A	90	11-1	6-7	7-11	16
B	97	12-0	7-5	8-1.5	8.5
C	92	12-2	7-6	9-0.5	18.5
D	85	11-7	7-9.5	8-3.5	6.0
E	92	11-8	8-0	8-7.5	7.5
F	96	11-9	8-7.5	9-8	12.5
Mean	92	11-8	7-8	8-7	11.5
Non Therapy Group					
AA	82	10-5	7-1.5	7-10.5	9.0
BB	97	12-9	6-11	7-8	9.0
CC	86	12-4	7-6	8-10.5	16.5
DD	88	11-11	8-5	9-6.5	13.5
EE	93	12-1	8-7.5	8-9.5	2.0
FF	107	11-11	8-10	8-9.5	−.5
Mean	92.1	12-3	7-11	8-7	8.[illegible]

All the subjects were initially tested with the Wechsler Intelligence Scale for Children and the Gates Advanced Primary Reading Tests. They were then divided into two groups by the paired comparison method and were matched for age, IQ, and initial reading ability. While both groups received remedial reading instruction, one group was arbitrarily chosen to participate in non-directive or group centered therapy meetings which followed the principles established by Rogers, (*17*) Hobbs, (*12*) and others and were conducted by the author.

The subjects met as a group once each week for one hour. The therapist started the meetings with an initial structuring of the situation. The subjects were told that the opportunity to speak freely about their feelings, attitudes, and past and present experiences is usually helpful in making a better adjustment to their present situation and eventually to their homes. They were told that they were free to discuss any topic without restraints or value judgments imposed upon them by the therapist. The therapist's rôle was to reflect and clarify to the group, the ideas, feelings and behavior that occurred in the group meetings.

The group meetings and the remedial reading program were continued for six months. The twelve subjects were then retested with an alternate form of the Gates Advanced Primary Reading Tests. The 'before' and 'after' reading test results plus other pertinent data are presented in Table I. The final reading results indicate that the group which received group psychotherapy in addition to remedial reading instruction showed the greater improvement in reading. The non-therapy group showed a range of improvement in reading from −.5 to 16.5 months for a mean gain of 8.25 months. The group that received group psychotherapy ranged in reading improvement from 6.0 to 18.5 months for a mean gain of 11.5 months. This latter group gained 3.25 months or 39.4 per cent more than did the non-therapy group.

SUMMARY AND CONCLUSION

Many approaches are presently being used to correct reading disabilities. Many of these approaches seem as often to be successful as unsuccessful. The present study hypothesized that it is the improvement in emotional adjustment that occurs because of the psychotherapeutic relationship between the reading clinician and the disabled reader which is the common factor in most of the remedial reading techniques which are in use. To test this hypothesis, one of two matched groups of retarded readers participated in a program of group psychotherapy in addition to the regular remedial program. This group showed a 39.4 per cent greater improvement in reading than did the control group. It is therefore concluded that with the subjects studied the psychotherapeutic relationship was an important factor in the correction of reading disabilities.

BIBLIOGRAPHY

1) G. Arthur. *Tutoring as Therapy.* New York: The Commonwealth Fund, 1946.

2) V. M. Axline. "Non-directive therapy for poor readers." *Journal of Consult. Psychol.*, 2, 1947, pp 61–69.

3) C. C. Bennett, *An Inquiry into the Genesis of Poor Readers*, Teachers College Contributions to Education, Number 755. New York: Bureau of Publications, Teachers College, Columbia University, 1935.

4) E. A. Betts. "Teacher analysis of reading disabilities," *Elementary English Review*, XI, 1934.

5) R. E. Bills. "Non-directive play therapy with retarded readers." *J. of Consult. Psychol.* 2, 1950, pp 140–149.

6) P. J. Blanchard. "Psychoanalytic contributions to the problem of reading disabilities." *Psychoanal. Study of the Child.* Vol. II, 1946.

7) A. L. Edwards. *Experimental Designs in Psychological Research.* New York: Rinehart & Co., 1950 pp 182–185.

8) B. Fisher. "A psychologist's evaluation of teacher's reports and suggestions for their improvement." *Ed. Admin. and Superv.* 38, 1952 pp 175–179.

9) E. Gann. *Reading Difficulty and Personality Organization.* New York: Kings Crown Press, 1945.

10) A. I. Gates. "Failure in reading and social maladjustment," *Journal of the National Education Association*, 25, 1936.

11) —— *The Improvement of Reading.* New York: The Macmillan Co., 1947.

12) N. Hobbs, in C. R. Rogers. *Client Centered Therapy.* Boston: Houghton Mifflin Co., 1951.

13) R. S. Hardwick. "Types of reading disability," *Childhood Education*, 8, 1932, p. 425.

14) M. Klein. "A contribution to the theory of intellectual inhibition," *International Journal of Psychoanalysis*, XII, 1931.

15) W. P. Percival. *A Study of the Courses and Subjects of School Failure.* Doctor's Dissertation, Teachers College, Columbia University, New York, 1926.

16) R. S. Redmount. "Description and evaluation of a corrective program for reading disabilities," *Jour. of Educ. Psychol.*, 39, 1948.

17) C. R. Rogers. *Counseling and Psychotherapy.* Boston: Houghton Mifflin Co., 1942.

18) F. J. Schonell. *Backwardness in the Basic Subjects*, Edinburgh: Oliver & Boyd Ltd., 1942.

19) J. Strachey. "Some unconscious factors in reading," *The International Journal of Psychoanalysis*, XI, 1930.

20) S. H. Tulchin. "Emotional factors in reading disabilities in school children," *Journ. Educ. Psychol.*, 26, 1935, p. 444.

21) P. G. Vorhaus. "Non-reading as an expression of resistance." Rorschach Research Exchange, 10, 1946, pp 60–69.

22) P. Witty, and D. Kopel. *Reading and the Educative Process*, Boston: Ginn and Co., 1939, p. 231.

"THE ESSENTIAL HIGH SCHOOL CONTENT BATTERY" AS A PREDICTOR OF COLLEGE SUCCESS

MARIE P. DOLANSKY

Boston University

The predictive power of the *Essential High School Content Battery* was examined in three New England Universities; namely, the Universities of Vermont, New Hampshire, and Maine. In addition, an intelligence measure was used; the relationship between the predictor variables and the criterion, i.e., first semester grade-point averages, was expressed by means of multiple correlation coefficients which are reported herein.

The Essential High School Content Battery is a newly published test[1] which can be given from the end of the ninth grade through the end of the twelfth grade, and with beginning college freshmen. It covers the four basic areas of high-school instruction; namely, science, mathematics, social studies, and English. Test I. Mathematics, contains items measuring fundamental skills in computation, vocabulary and concepts, understanding of functional relationships, application of mathematics to life problems, interpretation of mathematical graphs, knowledge of mathematical facts and formulas, interpretation of data in tabular form and knowledge of important theorems. Test II. Science, contains items covering science information, using the concepts of science, and using the methods of science. Test III. Social Studies, measures acquaintance with contributions of famous Americans, understanding of current social and political problems, understanding of vocabulary of social studies, knowledge of civic information, growth of American democracy, knowledge and understanding of global geography, knowledge of contributions of world leaders, understanding of international relationships, knowledge of sequence of events in United States history and knowledge of world events. Test IV. English, contains items measuring reading for information, vocabulary, business definitions, use of references, literature acquaintance, language usage, capitalization and punctuation, and spelling. It can be assumed that the Battery contains some of the determinants of college success, or more specifically of success in the freshman year. Examination of the freshmen curricula in all three universities revealed that the required courses contain English in all instances and courses in one or more of the other major fields of high-school instruction. It was therefore concluded that the Battery could be regarded as a logically valid predictor of college success to the extent to which (1) the test scores reflect true differences in scholastic achievement, (2) the objectives of college instruction in the four areas measured by the Battery are the same or related to the objectives of high-school teaching, and (3) the objectives of college instruction are properly evaluated and their attainment appropriately represented by the conventional numerical grades.

---

[1] D. P. Harry, and W. N. Durost. *Essential High School Content Battery, Manual of Directions.* Yonkers-on-Hudson: World Book Co., 1951, p. 1.

The experimental sample.—Out of some 6000 high-school seniors who were tested with the Battery in Vermont, New Hampshire, and Maine in the spring of 1950, about 240 were found to have been enrolled as freshmen in the fall of 1950 in the institutions mentioned above. Thus, grade-point averages could be obtained for some forty-three students in University X, seventy-three in University Y, and for one hundred eighteen students in University Z.

Reliability analysis.—The accuracy of prediction is limited by the reliabilities of the measures through which the performance is manifested. In order to be able to estimate to what extent an imperfect correlation between test and criterion is due to lack of overlapping in function and the extent to which it is due to lack of precision of measurement, split-half reliability coefficients were computed for the four subtests of the Battery. For this reason, each subtest was divided into two halves and special care was taken to make them equivalent with respect to (1) type or content of each item, (2) Flanagan validity indices, and (3) difficulty indices based upon a random sample of the New England population tested. Table I gives the results of the reliability analysis.

The above data are based on the total experimental sample of 234 college students enrolled in the three universities. On the whole the measures can be regarded as possessing satisfactory reliability for interpretation of group results. The reliability coefficients found in this study are generally lower than the ones reported in the

Manual.[2] This result was to be expected since the data are based on a relatively homogeneous sample, i.e., high-school students who had entered college. Thus the following validity data must be interpreted in the light of the somewhat lower reliability coefficients.

TABLE I.—SPLIT-HALF RELIABILITY COEFFICIENTS, MEANS AND STANDARD DEVIATIONS OF THE TWO HALVES OF EACH SUBTEST OF THE BATTERY

Test	r_{ab}*	r_{2a2b}†	M_a	SD_a	M_b	SD_b	N
Mathematics	0.82	0.90	16.56	5.97	16.06	6.09	240
Science	0.72	0.84	22.12	4.92	22.66	4.79	240
Social Studies	0.80	0.89	26.51	6.62	25.05	6.75	240
English	0.81	0.90	97.27	9.51	95.43	9.68	240

* Obtained Split-half reliability coefficient.
† Corrected by Spearman-Brown formula.

TABLE II.—PRODUCT-MOMENT CORRELATION COEFFICIENTS ESTIMATING THE RELATIONSHIP BETWEEN THE FOUR SUBTESTS OF THE ESSENTIAL HIGH-SCHOOL CONTENT BATTERY AND GRADE-POINT AVERAGES IN UNIVERSITIES X, Y, AND Z

Variable	University X (N = 43) r_{xy}	University Y (N = 73) r_{xy}	University Z (N = 118) r_{xy}
Mathematics Test	0.18	0.51	0.45
Science Test	0.31	0.57	[illegible]
Social Studies Test	0.40	0.50	[illegible]
English Test	0.57	0.44	[illegible]

Validity analysis.—In this analysis, each subtest scores were correlated with the criterion scores. This was done separately for the three schools involved. Table II gives the results. It can be seen that some of the correlation coefficients vary considerably from school to school. The results obtained in Universities Y and Z are more comparable than the ones in University X. The smallness of the sample in University X might have caused this effect. Other variations are most probably caused by differences in the variability of the groups, and by differences in the curricular offerings in the various schools. In any case the results demonstrate that a test's validity coefficient is subject to considerable changes with changes in the sample, or in the composition of a complex criterion.

---

[2] *Op. cit.* p. 15.

TABLE III.—PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN THE TERMAN-MCNEMAR TEST OF MENTAL ABILITY IQ'S AND GRADE-POINT AVERAGES EARNED IN UNIVERSITIES X, Y, AND Z

University	Correlation Coefficient	Mean IQ	SD of IQ's
University X	0.44	116.36	12.56
University Y	0.45	116.60	11.06
University Z	0.40	114.12	9.88

TABLE IV.—INTERCORRELATIONS AMONG THE FOUR SUBTESTS OF THE BATTERY AND THE TERMAN-MCNEMAR IQ'S

	Mathematics	Science	Social Studies	English	IQ's
Mathematics		0.58	0.48	0.43	0.55
Science			0.65	0.52	0.65
Social Studies				0.54	0.64
English					0.63
IQ's					

TABLE V.—BETA-WEIGHTS OF THE FIVE PREDICTOR VARIABLES COMPUTED SEPARATELY BY SCHOOLS

Test	University X	University Y	University Z
Mathematics	−.16	.24	.21
Science	−.05	.30	.32
Social Studies	.14	.16	−.01
English	.49	.14	.21
Intelligence Measure	.16	−.06	−.05

Since the Terman-McNemar Test of Mental Ability IQ's were available for most of the students in the experimental sample, their relationship to the grade-point averages was also examined. The results are reported in Table III. The average mental ability (as measured by this test) of the students in University Z appears to be slightly lower and the sample seems to be slightly more homogeneous. Students in University X and University Y were practically identical with respect to average mental ability. However, the above data must be viewed as pertaining to the experimental sample only, since not all freshmen in any of the three schools were tested, and sampling biases of various kinds might have been present. The predictive power of the IQ's was only moderately high and, as will be shown later, it did not contribute to the total prediction, since the test did not seem to measure any abilities different from those measured by the Battery itself.

In order to be able to combine the measures and compute a multiple correlation coefficient, intercorrelations between the five predictor variables had to be obtained. The data given in Table IV are based on the total of two hundred thirty-four cases. An examination of this table would suggest that the degree of independence among the variables tested was such as to make it reasonable to suppose that each would contribute something to the multiple prediction. However, one must also remember that the intercorrelations are in most instances as high or higher than the validity coefficients given in Table II. Thus, the multiple R can be expected to be only moderately high and the Beta-weights for many of the variables rather low. The obtained multiple R's, are in agreement with the above consideration. They equal to 0.60, 0.63, and 0.57 for Universities X, Y, and Z, respectively.

The Beta-weights for the various predictor variables, arrived at by the Doolittle solution of the normal regression equations are given in Table V.

The reliability of the criterion variable was not investigated in this study. It is very probable that it was rather low. Thus, it can be assumed that, if the obtained r's had been corrected accordingly, the multiple prediction might have been considerably improved.

Nevertheless, it can be said that the multiple R's obtained in this study compare favorably with most R's reported in studies of a similar nature. However, one must remember that an R of 0.60 gives one only about twenty per cent improvement over pure chance prediction. There is still room for much refinement in the general field of college success prediction.

MENTAL HEALTH ANALYSIS OF SOCIALLY OVER-ACCEPTED, SOCIALLY UNDER-ACCEPTED, OVERAGE AND UNDERAGE PUPILS IN THE SIXTH GRADE

VAGHARSH HAGOP BEDOIAN

North Hollywood High School

In recent years considerable attention has been directed to the study of mental health. In fact our Federal Government has considered mental health important enough to pass a Mental Health Act. Citizens have organized local, state and national mental health associations. Unfortunately the bulk of the energy has been applied to the adult population. The present study is concerned with the school population where, in many cases, the problems of the adult are rooted. Mental health problems of children were studied by seeking answers to the following questions: Is there any relationship between mental health, overageness and underageness? Furthermore, is there any relationship between mental health, social acceptance and social rejection?

Mental health was determined by administering Thorpe, Clark and Tiegs' *Mental Health Analysis*—Elementary Series, Form A to seven hundred and forty-three sixth-grade pupils in and around Los Angeles. This paper and pencil questionnaire differs from the usual personality test in the fact that its results can be summarized into two scores: Mental Health Liabilities and Mental Health Assets. The total score is obtained by combining the two raw scores of liabilities and assets and finding the percentile rank on a chart provided for that purpose.

Mental Health Liabilities include: A) Behavioral Immaturity, B) Emotional Instability, C) Feelings of Inadequacy, D) Physical Defects and E) Nervous Manifestations. The scoring is so arranged that a high percentile rank always indicates freedom from liabilities. Mental Health Assets include: A) Close Personal Relationships, B) Inter-personal Skills, C) Social Participation, D) Satisfying Work, and Recreation and E) Adequate Outlook and Goals. A high score on this group of subtests is intended to indicate a satisfactory and desirable amount of these manifestations of the child's personality.

Underageness and overageness were determined by taking those who were nine months below or above the mean chronological age of their half-grade placement. In twelve out of twenty-two classes the underage pupils had higher Mental Health Liability mean percentile scores than their fellow students. In six of the remaining ten classes the at-age pupils achieved higher mean scores and in only four classes the overage pupils earned higher mean percentile scores. The difference between the Mental Health Liabilities mean percentile scores of the underage[1] and overage pupils was 19.9 with a critical ratio of 4.6, which is significant beyond the one per cent level of confidence. The difference between the at-age and overage pupil's mean scores was 16.7 with a critical ratio of 4.3, which is significant beyond the one per cent level. The mean difference between the underage and at-age pupils' Mental Health Liabilities mean scores was not significant.

In thirteen out of twenty-two classes the underage pupils possessed higher Mental Health Assets mean percentile scores than the at-age and overage pupils. In four of the nine remaining classes the at-age pupils make higher mean percentile scores and the overage pupils received higher mean scores in the remaining five classes. The difference between the underage and overage pupils' Mental Health Assets mean percentile scores was 16.4 with a critical ratio of 3.7, which is significant beyond the one per cent level of confidence. The mean difference of the at-age and overage pupils scores was 13.6 with a critical ratio of 3.2, which is significant beyond the one per cent level. The underage and at-age pupils' mean difference was not significant.

In thirteen out of twenty-two classes the underage pupils earned greater Mental Health total scores. The at-age pupils received the highest Mental Health total scores in seven classes and the overage pupils in but two classes. The difference between the underage and overage pupils' Mental Health total scores was 21.1 with a critical ratio of 7.2, which is significant beyond the one per cent level. The at-age and overage pupils' mean difference was 16.3 with a critical ratio of 4.5, which is significant beyond the one per cent level of confidence. The mean difference between the underage and at-age pupils' Mental Health total scores was significant at the five per cent level.

---

[1] In presenting the differences of means, the first mentioned is always the greater of the two. This scheme is followed throughout the study.

Again one finds the stepped relationship among the underage, at-age and overage pupils. The younger pupils seem to have better self-evaluation than the others in the class. The at-age pupils received the next best self-evaluation and the overage pupils appear to have the lowest self-evaluation.

MENTAL HEALTH OF OVER-ACCEPTED AND UNDER-ACCEPTED PUPILS

Social acceptance and social rejection were determined by a multi-criteria sociometric test. Five questions were asked in the sociometric questionnaire: (1) If you were changed to another classroom, which boy or girl from this classroom would you choose to go with you? (2) If you were changed to another classroom, which boy or girl from this class would you not choose to go with you? (3) Which boy or girl from this classroom would you choose as a team captain in a game? (4) Which boy or girl from this classroom would you choose for class president? (5) If you needed help to do your classwork, which boy or girl from this classroom would you choose?

Each pupil was allowed three choices for each question. The choices were weighted as follows: five points for a first choice, three points for a second choice, and one point for a third choice. Sociometric status was determined by the combined weighted scores of question one, three, four and five. Standard scores were computed from the raw scores using Milo Whitson's[2] dual-deviation technique adapted for sociometric data. Significance above and below the mean were also computed. Those who were roughly three standard deviations above the mean were said to be significantly over-accepted. Those who were roughly three standard deviations below the mean were said to be significantly under-accepted. Rejection scores were compiled from question number two, ranked and the four highest scores taken. In nineteen out of twenty-two classes over-accepted pupils had higher Mental Health Liabilities mean percentile scores than the under-accepted pupils. The difference between the means was 23.6 with a critical ratio of 5.5, which is significant beyond the one per cent level. Not only are the over-accepted pupils more popular than their peers but they enjoy better mental health; furthermore the under-accepted pupils are less popular and by their evaluation have poor mental health.

---

[2] Milo Whitson, "Statistical and Geometric Techniques for Sociometric Data." Unpublished Doctor's Dissertation, University of Southern California, Los Angeles, Calif., June, 1949. 233 pp.

In Mental Health Assets the over-accepted pupils excelled over the under-accepted pupils in all twenty-two classes. The mean percentile scores of the over-accepted pupils were larger for each class studied. The difference between the means was 28.1 with a critical ratio of 5.2, which is significant beyond the one per cent level.

Data for the Total Mental Health scores reveal the same relationship with the over-accepted pupil enjoying better mental health in twenty-one out of twenty-two classes. The under-accepted pupils had higher mean percentile scores in only one class and that by a difference in score of only .80. The difference between the means was 27.5 with a critical ratio of 7.4, which is significant beyond the one per cent level.

The pattern is clear. The over-accepted pupils are better adjusted individuals as well as being better liked by their peers. It is difficult to say if there is a cause-and-effect relationship present. Determining validity for a published personality test is outside the scope of this study, but the data are so one-sided that it is difficult to ignore the apparent correlation between a high score on the *Mental Health Analysis* test and a high score on the sociometric questionnaire. The similarity becomes even more outstanding when one considers the great differences in the two types of questionnaires.

For a more complete study of the *Mental Health Analysis* questionnaire, the mental health of the four highest 'rejectees', the four highest 'stars of attraction' and the remainder of the group were compared.

MENTAL HEALTH OF 'STARS', 'REJECTEES' AND THE REMAINDER OF THE GROUP

The 'stars of attraction' excelled over the other two groups in Mental Health Liabilities in sixteen out of twenty-two classes studied. The remainder of the group received higher scores in four classes and the 'rejectees' in two classes. The mean difference between the 'stars' and 'rejectees' Mental Health Liabilities scores was 22.6 with a critical ratio of 5.4, which is significant beyond the one per cent level. The difference between the 'stars' and the remainder of the group was 7.8 with a critical ratio of 2.5, which is almost significant at the one per cent level. The mean difference between the remainder of the group and the 'rejectees' was 14.8 with a critical ratio of 4.6, which is significant beyond the one per cent level.

The 'stars of attraction' possessed the highest mean percentile score in all but one class in Mental Health Assets. In one class the remainder of the group received the highest mean percentile score. The difference between the means of the 'stars' and 'rejectees' was 30 with a critical ratio of 8.1, which is significant beyond the one per cent level. The 'stars' and the remainder of the group had a mean difference of 13 with a critical ratio of 5.0, which is significant beyond the one per cent level. The difference between the remainder of the group and the 'rejectees' was 17 with a critical ratio of 5.7, which is significant beyond the one per cent level.

The four highest 'stars of attraction' also lead in Mental Health total scores in sixteen out of twenty-two classes. The remainder of the group received the highest mean scores in five classes and the 'rejectees' Mental Health total scores was 28.4 with a critical ratio of 7.7, which is significant beyond the one per cent level. The mean difference between the 'stars' and the remainder of the group was 10.4 with a critical ratio of 3.7, which is significant beyond the one per cent level. The Mental Health total score mean difference between the remainder of the group and the four highest 'rejectees' was 18 with a critical ratio of 6.2, which is significant beyond the one per cent level.

SUMMARY

Mental Health Analysis data of the underage, at-age and overage pupils were presented in this study. Similar data were presented for the socially over-accepted and under-accepted pupils.

1) The underage and at-age pupils earned significantly better mental health scores than the overage pupils. The mean difference between the underage and the at-age pupils was not significant. In classes twelve and seventeen, where the overage pupils enjoyed better sociometric status, they rated themselves with average mental health. On the other hand, in classes four and eleven, where the overage pupils received relatively superior sociometric status, they rated themselves above average mental health. Whether the low mental health of the overage pupils is a real difference revealed by the test scores or an accidental one because overage pupils are usually found to be poor readers and do not understand test items sufficiently well to answer the questionnaire intelligently is an hypothesis worth investigating.

2) The socially over-accepted pupils produced significantly

higher mental health scores than the under-accepted pupils. The pupils who possess superior sociometric status also have better mental health than the pupils who are ignored, unwanted and disliked by their peers.

3) The four highest 'stars of attraction', those receiving the largest number of first and second choices on the sociometric test, rated themselves significantly higher mental health scores than the four highest 'rejectees' or those who received the greatest number of rejection choices on the sociometric test. The 'stars' also appear to possess better mental health than the remainder of the class.

BIBLIOGRAPHY

Robert B. Baron. *Mental Health Factors Associated with Social Acceptance and Social Rejection*. Unpublished Doctor's Dissertation, University of Southern California, Los Angeles, 1948, 173 pp.

Vagharsh H. Bedoian. *Sociometric Status and Mental Health of Overage and Underage Pupils in the Sixth Grade*. Unpublished Doctors Dissertation, University of Southern California, Los Angeles, 1950. 169 pp.

Emil Heintz. "Adjustment problems of class status," *Phi Delta Kappan*, 30: 290–293, April, 1949.

Raymond G. Kuhlen, and Howard S. Bretschi. "Sociometric status and personal problems of adolescents," *Sociometry*, 10: 122–132, May, 1947.

—— and Betrice Lee. "Personality characteristics and social acceptability in adolescence," *Journal of Educational Psychology*, 34: 321–340, September, 1943.

E. W. McElwee. "A comparison of personality traits of 300 accelerated, normal, and retarded children," *Journal of Educational Research*, 26: 31–34, September, 1932.

Jacob L. Moreno. *Who Shall Survive?* Washington, D. C.: Nervous and Mental Disease Publishing Co., 1934. 450 pp.

Mary L. Northway. "Outsiders: A study of the personality factors of children least acceptable to their agemates," *Sociometry*, 7: 10–25, February, 1944.

Adolph Sandin. *Social and Emotional Adjustment of Regularly Promoted and Non-Promoted Pupils*. Child Development Monographs, No. 32. New York: Bureau of Publications, Teachers' College, Columbia University, 1944. 142 pp.

Milo E. Whitson. *Statistical and Geometric Techniques for Sociometric Data*. Unpublished Doctor's Dissertation, University of Southern California, Los Angeles, June, 1941. 233 pp.

THE BLOOD PICTURE IN READING FAILURES

THOMAS H. EAMES

School of Education, Boston University

The observation that reading failures with anemia often improve in their school achievement without special teaching when the anemia is adequately treated, directed attention to the circulatory area. Two previous studies (*1*, *2*) showed a higher incidence of circulatory difficulties among poor readers than among non-failures or unselected pupils. A study made in connection with stuttering (*3*) (another aspect of language trouble) showed biochemical changes in the blood of stutterers. The present investigation concerns itself with hemoglobin and the blood cells.

The blood of thirty reading failures were studied as to the level of hemoglobin, the total red and white cell counts, and a differential white cell count. The results are summarized in Table I. The group ranged in age from seven through seventeen, with a median age of eleven. Eighty-three per cent of the failures were boys, approximately the usual proportion found in most language difficulty groups. The median birth weight of the group was seven and two-tenths pounds, but it was deemed significant that the first quartile point in the distribution fell at five and one-half pounds. This means that twenty-five per cent of the cases were children of premature birth, a considerably higher incidence than one meets in the general population. It tends to confirm an earlier finding that reading failures showed an incidence of premature birth (by either birth weight or time criteria) of fifteen per cent (*4*). About a third of the group exhibited hemoglobin of twelve and four-tenths grams or less, but the average and also the median hemoglobin for the whole group was thirteen grams, which falls in the normal range. The red blood cell count was not remarkable nor was the total white count.

The differential white count showed some variations though not statistically great ones. The polymorphonuclear leukocytes in both average and median for the group fell slightly below the control norm, which was the average of values for corresponding ages, taken from the work of several investigators. However, it may be significant that nearly three-quarters of the group fell below this norm when, statistically, one would expect not more than half to do so. The lymphocytes were slightly more numerous among the reading failures and the monocytes were slightly less frequent. Somewhat more than three-quarters of the cases fell below the monocyte norm. Eosinophiles and basophiles did not deviate very much from the adult level, although the basophiles were slightly more numerous.

TABLE I. BLOOD PICTURE IN THIRTY READING FAILURES

Characteristic	Reading failures				Norms.* Av.
	Q_1	Md	Q_3	Av.	
Hemoglobin (grams)	12.3	*13.0*	13.8	*13.0*	*13.5*
Red cells	4,450,000	*5,000,000*	5,420,000	*4,956,250*	*4,750,000*
White cells	7,000	*8,500*	10,320	*8,949*	*8,000*
Polymorphonuclear leukocytes	51	*57*	64	*54*	*62*
Lymphocytes	27	*36*	44	*34*	*30*
Monocytes	3	*6*	7	*5*	*8*
Eosinophiles	2	*3*	5	*4*	—
Basophiles	0	*1*	2	*1*	—
	Frequency	degree.			
Monocytes	*20%*	—			*0*
Myeloblasts	*0*	—			*0*
Lymphoblasts	*0*	—			*0*
Platelets	*10%*	*numerous*			*0*
Anisocytosis	*13%*	*moderate*			*0*
Poikilocytosis	13%	*slight*			0

* Norms in terms of average. Values from a number of different reports of studies of children in the same age range as this study.

A number of cases exhibited abnormal cell forms. Myelocytes were present in twenty per cent of the cases; platelets were numerous in ten per cent; while moderate anisocytosis with the presence of poikilocytes was observed in thirteen per cent. The total number of cases presenting abnormal forms constituted twenty per cent of the distribution. The forms observed suggest a tendency toward one or another form of anemia. No myeloblasts or lymphoblasts were found in any case.

This study is limited by the small number of cases, which is due to the fact that parents and pupils who come in for educational study and advice are not usually prepared for medical laboratory procedures. When a needle or other piece of medical equipment is mentioned or exhibited pupil-teacher rapport is promptly impaired and often destroyed. For this reason such work is avoided.

The results of this study suggest that, within the present group at least, blood cell and haemoglobin changes are not frequent enough to be of statistical importance and, as is often said in educational investigations, none of the factors involved will differentiate good from poor readers.

This implies that anemia is probably not a major cause of reading or other school subject failure. The fact, however, that more than a third of the pupils in the study showed low haemoglobin indicates the possibility that anemia may be a contributing cause in some instances. Haemoglobin is concerned with the distribution of oxygen to the tissues of the body. An optimum amount must be supplied if metabolism is to be normal. Probably the cells of the cerebral cortex are the most sensitive to oxygen deprivation. When haemoglobin is low there is not enough to supply the necessary oxygen to the tissues and certain changes occur. Among them are some degree of cortical depression, diminution of the acuity of vision and hearing, personality changes and impaired handwriting performance. Such factors can make it difficult for the pupil to apply himself sufficiently for satisfactory learning. It may be pointed out that many anemic people of all ages complain, among other things, of being tired and unable to concentrate.

The significance of the lower count of polymorphonuclear leukocytes is somewhat tempered by the fact that the count is known to be low in children. Nevertheless, the norm used is for children of the ages included in this study and this tends to demand somewhat more consideration of this factor than would be given otherwise. The presence of the various abnormal cell forms seems to this investigator to lend some added significance to anemia in the broadest sense, as one possible contributing cause of poor school achievement.

SUMMARY

Thirty reading failures were studied as to the blood picture. Certain variations in hemoglobin and cell count were found and a fifth of the cases presented abnormal cell forms. The findings are

discussed. While the changes in general are not frequent enough to be statistically significant, except for the appearance of abnormal cell forms, it is concluded that they merit attention in the individual case and that they may constitute one of many contributing causes of reading failure.

REFERENCES

1) T. H. Eames. "A frequency study of physical handicaps in reading disability and unselected groups." *Journal of Educational Research* 29: Sept., 1933, pp 1–5.

2) —— "Incidence of diseases among reading failures and non-failures." *Journal of Pediatrics* 33: November, 1948, pp 614–617.

3) G. A. Kopp. *Metabolic Studies of Stutterers, Part I. Biochemical Study of Blood Composition.* Speech Monographs, 1: 1934, pp 117–132.

4) T. H. Eames. "Comparison of children of premature and full-term birth who fail in reading." *Journal of Educational Research*, March, 1945. pp 506–8.

BOOK REVIEWS

DANIEL BROWER AND LAWRENCE E. ABT, Editors. *Progress in Clinical Psychology*. New York: Grune and Stratton, 1952. Vol. 1, Section 1, pp. 328; Section 2, pp. 329–564.

The two books are presented as a single continued volume with a subject index in each book. The subjects through the two books are in seven parts: An introduction on the emergence of clinical psychology; diagnostic and evaluative procedures; psychotherapy; developmental processes; applications of clinical psychology to special areas; approaches to clinical psychology; and professional issues.

The forty-two chapters (twenty in the first book) are written by different individuals all of whom have done special work in the fields about which they write. For the most part the selection of authors has been excellent and the reports can therefore be considered as having a high degree of authority.

In planning this volume it was decided "to employ value judgments in the selection of material to be covered and in the allotment of space to the areas decided upon. We have sought, in Volume 1, to provide as complete a coverage as possible of the past six years in clinical psychology and to point up, in the process, as many stimuli as possible to further thinking and research." (p. vii). It was planned that the book should not be exhaustive nor "encyclopedically complete," but that each writer should be "as selective and constructively critical of his materials as he wished." (p. viii). It is one of a series of volumes to be issued every second or third year.

The introductory chapter by L. E. Abt gives a frank and over-all orientation in the present condition of clinical psychology, emphasizing the lack of generally acceptable principles, the dubious state of many of the principal tools of clinical psychology, lack of agreement on such problems as psychodiagnosis, diagnosis and prediction, difficulties in basic theory, the effort to integrate clinical psychology with "the larger science of psychology," and problems of psychotherapy, especially, of a research basis for psychotherapy. What is not to be forgotten is "The very youthfulness" of clinical psychology.

The content of the first book includes diagnostic and evaluative procedures and psychotherapy. Under the first head are discussed intellective functions; measures of aptitude, achievement and interest; personal documents; self-appraisal methods; testing for psychological deficit; the Rorschach test; apperceptive methods; House-tree-person and human figure drawings; Gestalt functions; sentence completion and work association tests; picture-frustration; and the Szondi test. Under the second head are client-centered procedure; psychoanalytic theory and technique; group psychotherapy; play and related techniques; spontaneous art in therapy and diagnosis; and neurosis and its treatment as learning phenomena.

The second book deals with developmental processes: infancy; early childhood; latency period; adolescence; and gerontology;—applications of clinical psychology to eleven special areas, (eleven chapters); approaches to clinical psychology: psychosurgery, cultural anthropology; social psychology, statistical methods; and P-technique factorization;—and clinical psychology as a profession.

The discussions for the most part are very condensed; so much so that a chapter, for example, of six and a quarter pages has a bibliography of 394 references. Many of the chapters have bibliographies running from nearly one hundred to more than two hundred references. The books are therefore especially valuable for reference.

The discussions assume a very considerable fundamental knowledge of the subject. The books do not attempt to give the fundamentals but explicitly state that they are concerned with the progress of the subject since World War 2.

Besides the bibliographies, the especial value of the books lies in the mature judgment of most of the writers who were chosen for their special knowledge and experience.

If there appear to be defects of balance in the contents and amounts of space given to different subjects, it is to be remembered that there were differences in the amounts of progress in the various fields of a new and fast developing discipline. It is too early to evaluate progress in the field, but these and later volumes which are promised may well provide material that will assist in such a valuation and which will point out holes in development that will attract research students to investigation that is most needed. Well may these books help emphasize how much we have not accomplished, as for example, in understanding, and, if possible, more agreement, in the problems of diagnosis and of psychotherapy. One might expect that some reference would have been made to the extensive work that has been done on professional ethics. Georgia has not certification but licensing. More detailed indexes would have added to the value of the books. The term diagnosis does not appear in either index.

In this field where there is great need for satisfactory texts these books are most welcome, but for students without a very considerable foundation, they will be found to be difficult and not wholly effective as texts.

A. S. EDWARDS
The University of Georgia

SAMUEL J. BECK. *Rorschach's Test II—Advances in Interpretation.*
New York: Grune & Stratton, 1952, pp. 301. $5.50.

In this volume, as in his earlier two, Dr. Beck continues his exhaustive analysis of the Rorschach Test and its ever widening compass. Like the first two volumes, it is characterized by a scholarship and a literary flavor that has come to be associated with Beck's writings; even more than the first two, it will serve the experienced Rorschacher rather than the beginner.

Rorschach's Test III is composed of three main sections. The first, and shortest, is a defense of the holistic concept of personality in which Beck seeks to combine the points of view of Hughlings Jackson, Freud, Lewin and the Gestalt school, with special emphasis on Jackson's concept of levels of functioning. Beck rejects in particular Allport's principle of independent traits and the view that the continuity of the individual is historic rather than functional. This uncompromising holism is not implicit in Rorschach's original approach (Beck would disagree), but it is necessary to support Beck's conviction that the entire personality is describable by the symbols of the Rorschach Test and its quantitative ratios. This may represent the experience of psychologists whose test armamentarium is restricted to the Rorschach, but the claim would be widely questioned by clinicians using other instruments along with the Rorschach in evaluating personality. There is no test so good but that another one may be more revealing at different times or for different individuals. Human personality is too complex to project itself in its totality on a single surface.

The second section, entitled "Advances in Interpretation," comprises not so much new material as further exploration and reinterpretation of the conventional Rorschach symbols. Of particular interest is Beck's discussion of the Rorschach elements which are claimed to reveal ego-strength, ego-insufficiences, and ego-defenses. The items particularly emphasized are the F+ %, the amount of W and Z, the sequence and approach, and interestingly enough, the P %—the determinants which Beck associates with intellectual productivity in contrast to those which deal with manifestations of emotion (C, Y, V, M). He regards P as a special and most concentrated form of the F+ response. According to Beck, a subject's P per cent indicates not only closeness of identification with the norm setting group but reflects further the degree of the subject's ability to integrate successfully. Also re-evaluated are the various affective determinants of the Rorschach, and in particular the M response to which Beck has always attached great significance. He is interested especially in M as a measure of energy, that is, the amount of energy invested by the patient in the M associations. In this connection Beck tried out the Levy Movement Scale and applied its scoring scheme to the movement responses on the Rorschach. His aim was to get some indication of how strongly the patients felt the indicated fantasies, but he was unable to get any definitive answer from the clinical material. Altogether this chapter will be found to be very provocative but, although reference is made to studies either in progress or about to be published, few factual data are presented.

In the third, and by far the longest section, Beck devotes over two hundred pages to detailed analyses of seven Rorschach protocols obtained from four patients, three of whom were tested before and again after therapy of varying duration. In these analyses Beck gives renewed evidence as to why he is estimated by many as one of the leading, if not the leading, Rorschacher in the country. To be able to write forty pages on a single protocol involving some thirty responses is in itself *tour de force*; to be able to do so and be convincing would be a greater achievement. In this respect Beck is only partially successful. The reports are much too long for even an initiated Rorschacher to follow. They are made additionally difficult by Beck's penchant for metaphor and analogy. The net result is that what he has to say is very interesting at any given point but far from clear when taken as a whole; nor do the summaries at the end of each case help much to clarify the total picture. As always, Beck's case presentations are profitable reading, but one hopes that in his next publication he will be a little less literary and more succinct. DAVID WECHSLER

Bellevue Hospital
New York University

DAEL WOLFLE, CLAUDE E. BUXTON, CHARLES N. COFER, JOHN W. GUSTAD, ROBERT B. MACLEOD, AND WILBERT J. MCKEACHIE. *Improving Undergraduate Instruction in Psychology*. New Pork: The Macmillan Company, 1952, pp. 60. $1.25.

As the authors state, an occasional review of what is being taught in a field is required to improve teaching a particular subject matter. The 'reviewing' was done by the authors at a summer conference. To many readers the results of their conferring will not only be found disappointing, but also will appear to be a backward rather than a forward step in teaching undergraduate psychology. Apparently their view is that undergraudate psychology is for the chosen few who may be interested and able to profit by a systematic presentation of scientific content in a narrow, systematic theoretical framework. No consideration is to be given to adjustment problems or materials related to contemporary life affairs. If student enrollment drops off markedly, so much the better. Staff members without courses to teach, they state, can be absorbed by laboratories.

The six chapters deal with objectives, recommended curriculum, personal adjustment courses, technical training, implementation of the curriculum, and research problems. In general, the viewpoint taken by the authors will seem to many to be narrow, one-sided and inadequate. There should be satisfactory alternatives to the program of the authors. In fact there are such programs already in existence and they are highly successful.

Finally, one may note that the title of this book is misleading. It deals mainly with programs of offerings and content of courses. Only incidentally is note made of instruction. Actually the contents of the volume might be described as a program of psychology for the psychologist rather than psychology for students.

University of Minnesota MILES A. TINKER

ARTHUR E. TRAXLER, ROBERT JACOBS, MARGARET SELOVER AND AGATHA TOWNSEND. *Intoduction To Testing and the Use of Test Results in Public Schools* New York: Harper and Brothers, 1953, pp. 113.

The increasing large scale use of objective tests by public school systems throughout this country has created a demand for a simplified but relatively complete handbook on the fundamentals of educational testing. The work of Traxler, et al is an attempt to meet this need. It is written in a way that any teacher, despite his lack of knowledge of statistics or psychometrics, can readily grasp the basic concepts which are presented and which are a fair sample of those prerequisite to any intelligent use of current measuring devices in the classroom.

The book is intensely practical in its orientation, and it seems apparent that the authors have had firsthand experience with classroom teachers of many types. The principles of selecting, administering, scoring, analyzing, recording and using test results which are presented are undoubtedly sound, and the brevity, clarity, and conciseness with which they are presented will appeal to the busy educator already overloaded with a full teaching or administrative schedule. The final chapter in the book particularly will find hearty acceptance even among teachers who are generally familiar with the contents of the preceding nine chapters. Here is presented the case history of an incoming student along with a detailed account of what tests were given him and when, why these particular tests were selected, how the scores were interpreted and used, and what the end-result of the program was upon his graduation.

This reviewer, however, believes that the naïve reader should be made aware of the fact that the book represents a somewhat narrow philosophy of educational measurement. Tests are conceived of as tools which the teacher can employ as aids to guide the learning experiences of students. Practically no reference is made to the fact that tests can serve other equally worthy purposes. The use of measurement as an active rather than a passive agent in the learning process receives no emphasis. The use of measurement as a powerful method for inducing and directing curriculum change is barely touched upon. The notion that students, as well as teachers, might profitably have some hand in the selection of the tests which are to be used for their own guidance, is entirely absent.

Although such criticisms may seem of relatively minor importance, there is some danger that users will fail to realize the fact that the book presents only one of the important points of view about educational tests and their uses. For this particular point of view, the book is quite good and will undoubtedly find wide acceptance among persons not intimately acquainted with tests.

Dora E. Damrin

Educational Testing Service Princeton, N. J.

James Deese. *The Psychology of Learning*. New York: McGraw-Hill Book Company, 1952, pp. 398. $5.50.

This undergraduate text, with an emphasis upon experimental evidence rather than upon theoretical points of view, is intended to provide a representative picture of the psychology of learning. The various chapters cover the important topics in the field. Special stress is placed upon recent literature.

The author states that "the subject matter of this book is the behavior of organisms." It is behavioristic "chiefly in the sense that it emphasizes observable animal or human behavior." It is intended that the theoretical framework of the book have wide application. To achieve this a great many examples of learning are covered with a relatively economical set of principles and concepts. The assumption is made that much of behavior, whether animal or human, is made of the same stuff. However, although many examples of learning are based upon rat behavior, the ultimate interest is in man.

This text is concerned with basic factors and principles in learning behavior. Included are various sections dealing with reinforcement, complex learning, retention, transfer, efficiency, problem-solving, nature of learner, conflict, physiological problems, and theory.

Instructors who are looking for an applied psychology of learning will not be interested in this text. But those who wish a brief straight-forward discussion of fundamentals will find that this book fits. Furthermore it is clearly written, well organized, and well documented.

Miles A. Tinker

University of Minnesota

PUBLICATIONS RECEIVED

David Abrahamsen. *Who Are the Guilty? A Study of Education and Crime.* New York: Rinehart & Co., Inc., 1952, pp. 340. $5.00.

Virgil A. Anderson. *Improving the Child's Speech.* New York: Oxford University Press, 1953, pp. 333. $5.50.

Melbourne S. Applegate. *Understanding that Boy of Yours.* Washington, D. C. (2153 Florida Ave.): Public Affairs Press, 1953, pp. 52. (paper)

Roger G. Barker, Beatrice A. Wright, Lee Meyerson, and Mollie R. Gonick. *Adjustment to Physical Handicap and Illness: A Survey of the Social Psychology of Physique and Disability.* New York: Social Science Research Council, 1953, pp. 440. (paper) $2.00.

Franziska Baumgarten. *Lehrling und Lehre.* Zurich, Switzerland: Rascher Verlag Zurich, 1952, pp. 238. (paper)

Robert H. Beck, Walter W. Cook, and Nolan C. Kearney. *Curriculum in the Modern Elementary School.* New York: Prentice-Hall, Inc., 1953, pp. 584. $4.95.

Samuel J. Beck. *Rorschach's Test. III: Advances in Interpretation.* New York: Grune & Stratton, 1952, pp. 301. $5.50.

Leopold Bellak, Editor. *Psychology of Physical Illness.* New York: Grune & Stratton, 1952, pp. 243. $5.50.

George K. Bennett, Harold G. Seashore, and Alexander G. Wesman. *A Manual for the Differential Aptitude Tests.* New York: The Psychological Corporation, 1952, pp. 77. (paper)

M. E. Bennett. *College and Life: Problems of Self-discovery and Self-direction.* Fourth Edition. New York: McGraw-Hill Book Co., 1952, pp. 457. $4.50.

Harold W. Bernard. *Mental Hygiene for Classroom Teachers.* New York: McGraw-Hill Book Co., 1952, pp. 472. $4.75.

Karl S. Bernhardt. *Practical Psychology.* Second Edition. New York: McGraw-Hill Book Co., 1953, pp. 337. $3.75.

F. K. Berrien. *Practical Psychology.* Revised Edition. New York: The Macmillan Co., 1952, pp. 640. $5.00.

Howard R. Bierman and Others. *Cancer Learning in the Medical School.* Berkeley, Calif.: University of California Press, 1952, pp. 87.

HUBERT BONNER. *Social Psychology: An Interdisciplinary Approach.* New York: American Book Co., 1953, pp. 439. $4.25.

HENRY BOWERS. *Research in the Training of Teachers.* Toronto: The Macmillan Co. of Canada, 1952, pp. 167. $1.90.

DANIEL BROWER AND LAWRENCE E. ABT. *Progress in Clinical Psychology.* Volume 1, Section 1. pp. 328. $5.75. Volume 1, Section 2, pp. 236. $5.00. New York: Grune & Stratton, Inc., 1952.

MARGARET H. BULLEY. *Art and Everyman: A Basis for Appreciation.* Two volumes. London: B. T. Batsford, Ltd., 1951, 1952, I: pp. 75 # 341 Illustrations; II: pp. 91 # 497 Illustrations. $12.75. Single volumes $6.50 each.

ERNEST W. BURGESS AND PAUL WALLIN. *Engagement and Marriage.* Philadelphia: J. B. Lippincott Co., 1953, pp. 819.

OSCAR KRISEN BUROS, Editor. *The Fourth Mental Measurement Yearbook.* Highland Park, N. J.: The Gryphon Press, 1953, pp. 1163. $18.00.

WILLIAM H. CARTWRIGHT AND WILLIAM B. HAMILTON, Editors. *The Duke University Centennial Conference on Teacher-training.* Durham, N. C.: Duke University Press, 1953, pp. 119. (paper)

LAWRENCE E. COLE. *Human Behavior: Psychology as a Bio-social Science.* Yonkers-on-Hudson, N. Y.: World Book Co., 1953, pp. 884. $5.50.

JAMES B. CONANT, Chairman. *Education for All American Youth: A Further Look.* Washington, D. C.: Educational Policies Commission, 1952, pp. 402. (paper) $2.00.

WALTER W. COOK AND WM. CLARK TROW, Co-chairmen. *Educational Psychology in Teacher Education: A Guide for Instructors.* Ann Arbor, Mich.: National Society of College Teachers of Education, 1953, pp. 99. (paper) $1.50.

STEPHEN M. COREY. *Action Research to Improve School Practices.* New York: Bureau of Publications, Teachers College, Columbia University, 1953, pp. 161. $3.00.

J. CORNWELL. *An Orally Presented Group Test of Intelligence for Juniors.* London: Methuen & Co., Ltd., 1952, pp. 40. 3s, 6d.

CLAUDE C. CRAWFORD, ETHEL G. COOLEY, C. C. TRILLINGHAM, AND EMERY STOOPS. *Living Your Life.* Second Edition. Boston: D. C. Heath & Co., 1953, pp. 448. $2.80.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 44 November, 1953 Number 7

AN EVALUATION OF SMALL GROUP WORK IN A LARGE CLASS

GOODWIN WATSON

Teachers College, Columbia University

The division of large classes into small groups for discussion has been practiced at Teachers College for more than thirty years. Every teacher who uses small discussion groups has found that some students enjoy the experience and rate it as more valuable to them than any other phase of the course. Others complain that their group got nowhere; that the discussions were a waste of time. Few efforts have been made to find out why the group experience means so much more to some students than to others.

This report covers reactions of three hundred and fifty graduate students taking a basic required course entitled Education as Personal Development. Each student worked in a small discussion group. There were forty-two groups, ranging in size from four to fifteen members; the average size was eight. A questionnaire and pre-test at the beginning of the course provide data which can be related to the appraisal each student made of his group experience.[1] On the final evaluation form each student rated his group experience for enjoyment, accomplishment, and for what he learned from his group. As shown in Table I, the most typical response was: Enjoyed the group; good group spirit; fine people; group accomplishment only about average, but I learned more in this group than in most courses. About two-thirds of the students gave this warm general endorsement. As compared with other class activities (Table II) the group discussions were thought to be about equal in value to the required readings; less useful than staff lectures, but more beneficial than films, panels, written work or supplementary reading.

Combining the several appraisals of small group discussions, [illegible] selected a 'high value' group of fifty-five students and a 'low value

---

[1] Analysis of the data was aided by a grant from the Faculty Research Fund of Teachers College.

TABLE I. COMPONENTS OF ATTITUDE TOWARD GROUP WORK

	Value Rating for Group Work			
	High value	Moderate value	Low value	All
N =	53	185	52	290
Enjoyment		*Percentage*		
Marvelous group; warm fellowship; high enjoyment	55	9	—	15
Enjoyed group; good group spirit; fine people	45	83	36	68
Enjoyed certain individuals, not so much the group as a whole	—	7	48	13
Group only fair; sessions dull; little group feeling	—	1	12	3
Did not enjoy	—	—	4	1
	100	100	100	100
Accomplishment				
Very proud; extraordinary	8	—	—	1
More than most	76	25	4	30
About average	16	69	56	58
Less than average	—	4	25	7
Nothing but talk	—	2	15	4
	100	100	100	100
Learning				
As educative as any experience I ever had	48	20	2	21
More than most courses	48	47	17	42
Like average course	4	27	40	25
Less than average course	—	5	35	10
Complete waste of time	—	1	6	2
	100	100	100	100
Rank for Value Among Class Procedures				
1 (Best)	77	10	—	21
2	17	25	2	19
3	6	20	—	14
4		24	6	16
5		16	23	14
6		4	33	9
7 (Poorest)		1	36	7
	100	100	100	100

group of fifty-one students. Because of the generally favorable attitude toward small group sessions, the 'low value' group necessarily includes some who, as shown in Table I, rated their group as enjoyable, its accomplishment as 'average' and their own learning as at least equal to that in an average course. On the whole, however, (Table II) the 'high value' students ranked their group experience higher than any other course activity while the 'low' value students ranked it below every other course activity. The rest of this report will be concerned with hypotheses designed to predict the 'high' and 'low' evaluations. First, we shall examine four hypotheses which concern the group as a whole, employing a 'molar' rather than 'molecular' concept of group value.

TABLE II. COMPARATIVE EVALUATION OF COURSE ACTIVITIES

	Mean Rank Assigned Each of Seven Kinds of Course Activity by Students Giving Group Work:			
	High value	Moderate value	Low value	All
N =	55	243	51	349
1) Staff lectures	2.80	2.25	1.61	2.23
2) Required readings	4.40	3.33	3.06	3.43
3) Group discussions	1.34	3.42	6.03	3.47
4) Moving pictures	4.70	4.19	4.02	4.23
5) Panel discussions	4.47	4.59	4.12	4.51
6) Writing professional diary	4.58	4.64	4.72	4.64
7) Voluntary, supplemental reading	5.59	4.94	4.37	4.95

Hypothesis 1. Groups are homogeneous in judging the value of their experience; good groups tend to be rated high by all members; poor groups tend to be rated low by all members; the differentiating factors must be sought not in personal characteristics of individuals but in some features common to the whole group.—Actually, there were no groups in which all members agreed in attributing either 'high' or 'low' value to their group experience. The hypothesis of homogeneity would lead us to expect 'high' rating members to be concentrated in perhaps seven or eight excellent groups; the 'low' rating members in an unfortunate seven or eight. Actually the fifty-five 'high' raters were distributed among twenty-four different groups; the fifty-one 'low' raters were found scattered throu

twenty-six different groups. The distribution of high's and low's throughout the groups did not differ significantly from what chance alone would have given. There were eleven groups in which one or more members gave a 'high' rating while one or more others gave the same group experience a 'low' rating.

The first hypothesis must be rejected. The students who get most out of their group work are not concentrated in certain outstanding groups, nor is a 'poor group' the main reason for low ratings. Value judgments by members in the same group vary almost as widely as do those of invididuals selected at random from various groups.

TABLE III. SIZE OF GROUP AND VALUE RATING

Size of Groups	No. of Groups	Value Rating*			
		High	Moderate	Low	All
4–6	12	6	41	5	52
7–9	15	11	82	28	121
10–15	15	36	109	19	164
	42	53	232	52	337

* Distributions significally different at .01 level.

Hypothesis 2. Average value rating will decrease as groups increase in size.—The hypothesis rests upon Bales' observation that as groups grow from four to fifteen members, there is increasing tendency for one member to dominate and for some to remain only listeners. Our smaller groups might therefore be expected to have more nearly equal participation by their members, and hence a higher level of satisfaction.

The data reported in Table III show a statistically significant relationship, but not in the expected direction. The large groups had, proportionately more 'high' ratings. In the one group as large as fifteen members, eight reported high satisfaction. One group of eleven members produced eight ratings of 'high', yet another group the same size produced five 'low' ratings and no 'high'. Groups with most low ratings included neither the largest nor the smallest groups, but were predominantly of average size; the smallest groups lded more moderate and fewer extreme appraisals.

The third molar hypothesis concerns two types of work carried on by these groups. A random half of the groups was arbitrarily directed to select a single problem of special interest to members and to work continuously on that subject all semester. The remaining groups were directed to discuss in their group each week some issue arising from the accompanying class session; hence they talked about a different question at each meeting.

Hypothesis 3. Groups working continuously and cumulatively on a single project will be experienced as more valuable than are groups discussing a variety of issues more superficially; the difference in their sense of accomplishment will be especially marked.—The results, presented in Table IV, show no advantage for either type of group, in enjoyment, accomplishment, learning, or in the composite value judgment.

A fourth molar hypothesis is that those students who find themselves in a group which is generally able and well-advanced will get more stimulation and will find more value in their group than will be the case for students who happen to draw a group made up of the less competent. One indication of competence is score made at beginning of the term on a test of thirty items central to the course content.

Hypothesis 4. Level of satisfaction with the group will vary with the group's average level of competence in course knowledge.— Actually 'high value' ratings came from members of groups which averaged 22.1 on the pre-test; 'moderate value' members belonged to groups averaging 22.4; while 'low value' members came from groups averaging 21.8. The differences are not statistically significant. The hypothesis cannot be defended with these findings. Whatever it was that made a group 'good' was not measured by the pre-test. Relation of satisfaction with group work to an individual's own pre-test score will be considered later. (Hypothesis 8)

We turn now to hypotheses concerning individual personality differences which may be expected to be associated with a liking for small group discussion. An obvious expectation would be that individuals (especially by the time they have become graduate students of education) themselves know whether they find group work profitable or not. Two hypotheses arise.

Hypothesis 5. Students who declare in advance that they "usually prefer working as a member of a coöperative group" will find the group experience more rewarding than will those who usually "pre[illegible] working individually".

TABLE IV. TYPE OF GROUP PROGRAM AND VALUE RATING*

	Single Term Project	Various Topics	Total
A. *Enjoyment*			
1) Marvelous group; warm fellowship; high enjoyment	28	36	64
2) Enjoyed group; good group spirit; fine people	106	115	221
3) Enjoyed certain individuals; not so much the group as a whole	22	20	42
4) Group only fair; sessions dull; little group feeling	3	7	10
5) Did not enjoy the group experience	2	1	3
	161	179	340
B. *Accomplishment*			
1) Very proud of extraordinary achievements	3	4	7
2) More than most other groups	49	49	98
3) About average	89	107	196
4) Less than average	12	13	25
5) Nothing but talk	6	7	13
	159	180	339
C. *Learning from group*			
1) As educative as any experience I ever had	34	34	68
2) Learned more than in most courses	69	73	142
3) Learned about as much as in average course	36	51	87
4) Learned less than in average course	17	17	34
5) A complete waste of time so far as new learning is concerned	4	2	6
	160	177	337
D. *Composite value judgment*			
Rated high in value	25	27	52
Rated moderate in value	111	76	187
Rated low in value	24	27	51
	160	130	290

* Distributions not significantly different

Hypothesis 6. Students who report in advance that they usually take a leading, active rôle in group discussion will place higher value upon group work than will those who are usually reticent and let others do the talking.

The data fail to support either hypothesis. The response "Prefer working as a member of a coöperating group" was given at the beginning of the course by seventy-two per cent of those who rated their group experience low in value, and by seventy-four per cent of those who rated it high. As shown in Table V those who claimed an active rôle in discussions were no better satisfied than were those who declared their participation to be average or usually deficient. Apparently we won't get far toward meeting student needs by following their advance expression of preference for group versus individual work.

TABLE V. USUAL RÔLE IN GROUP DISCUSSION AND VALUE RATING*

Usual rôle	Value Rating for Group Work			
	High	Moderate	Low	All
Active; take a lead	14	49	12	75
About average participation	37	166	35	238
Reticent; let others talk	2	26	4	32
	53	241	51	345

* Distributions not significantly different.

Further light on the weight to be given to preference expressed in advance, comes from analysis of student reactions to the two types of group procedure. They were asked in advance whether they preferred "a group which takes up a different problem each week touching on many areas during the term", or "a group which concentrates on one problem area for the entire term, giving a more thorough grasp of a limited topic." We have already reported that the two types of group proved equally satisfying, but the question here concerns the relation of advance preference to eventual satisfaction. Student preference was ignored in making the group assignments, an arbitrary procedure which aroused protest from a few students, and which was defended as necessary for the larger experimental design. Two hypotheses appear plausible [illegible] this connection:

Hypothesis 7. Students who prefer the wider variety of issues are more likely to favor group discussion; students who prefer to concentrate and do more thorough work on a single problem are less likely to be satisfied with a discussion group.

TABLE VI. TYPE OF GROUP PROCEDURE PREFERRED AND VALUE RATING GIVEN GROUP WORK*

Preference	Value Rating for Group Work			
	High	Moderate	Low	All
Strongly prefer various topics	22	97	14	133
Mildly prefer various topics	19	62	18	99
Don't care	0	9	2	11
Mildly prefer single project	10	37	5	52
Strongly prefer single project	4	34	10	48
	55	239	49	343

* Distributions not significantly different.

TABLE VII. ASSIGNMENT WITH OR AGAINST EXPRESSED PREFERENCE AND SUBSEQUENT VALUE RATING GIVEN GROUP WORK

Relation of Assignment to Previous Choice	Value Rating for Group Work			
	High	Moderate	Low	Total
In accord with strong preference	11	48	15	74
In accord with mild preference	18	42	16	76
Expressed no preference	0	9	2	11
Counter to mild preference	13	31	6	50
Counter to strong preference	10	51	13	74
	52	181	52	285

* Distributions not significantly different.

Hypothesis 8. Students assigned to the type of group they prefer will have a more satisfying experience than will students assigned in contradiction to their preference.

As reported in Table VI, there is only very weak support for the seventh hypothesis. A difference in the expected direction appears only in the categories of strong preference and the cases [illegible]re are so few that statistical confidence is unwarranted.

From Table VII it may be observed that the distribution of satisfaction among the seventy-four students assigned in accord with a strong preference is almost identical with the satisfaction distribution among those assigned counter to a strong preference. Hence Hypothesis 8 must be rejected.

Another source of data concerning individual personality characteristics is a check-list on which students, at the beginning of the term, registered the emphasis they wished to give to each of eight proposed course topics. Previous investigations have shown that some persons are prone to accept doubtful statements while others more often reject such statements or remain uncertain. It seemed possible that analogously there might be some students prone to accept and to approve whatever the teacher offered, and others generally negatively disposed. The former might be expected to respond to many proposed topics with enthusiasm and also to report favorably on their group experience. The negativists would similarly be skeptical of the value both of course topics and of group work. Hence:

Hypothesis 9. Group value ratings are positively correlated with ratings of course topics for anticipated value.—Before the data are presented, three other hypotheses will be introduced, since these depend upon the same preliminary check-list.

Hypothesis 10. Group-prone personalities will give higher advance rating to the topic, "Procedures in group leadership".

Hypothesis 11. Individualists have more difficulty in inter-personal relations; this will be expressed in higher advance ratings for: "Overcoming inferiority or inadequacy feelings"; "Fears and anxieties"; "Sex adjustment"; and "Psychotherapy".

Hypothesis 12. Students who express more interest in proposed topics such as techniques for discipline, testing and guidance, reveal a manipulative attitude toward others and may be expected to rate group work lower.

The data of Table VIII fail to support most of these hypotheses. Hypothesis 9 fails since, in the last two lines of the table it appears that the average student gave 3.2 single checks and 1.45 double-checks and that this holds true of all groups with no significant deviation.

The topic of "Procedures in group leadership" was ranked in sixth place by the group-prone persons and in fourth place by the individualists; a difference counter to our hypothesis but not statistically significant.

TABLE VIII. TOPICS PREFERRED FOR EMPHASIS IN COURSE

Topics (In order of preference)	Value Rating for Group Work							
	High		Moderate		Low		All	
	Av. score	Rank	Av. score	Rank	Av. score	Rank	Av. score	Rank
Discipline: how combine order and freedom in practical class situation	91	2	100	1	80	3	97	1
Techniques of guidance and counseling	87	4	89	2	92	1	89	2
Overcoming feelings of inferiority or inadequacy in inter-personal relations	89	3	88	3	63	6	85	3
Fears and anxieties; how they arise and are removed*	98	1	79	4	67	5	80	4
Procedures in group leadership	60	6	75	5	76	4	73	5
Value of psychotherapy for personal development	56	7½	72	6	88	2	72	6
Problems of sex adjustment in the modern world	69	5	64	7	57	8	64	7
Tests (intelligence, aptitude, interests, etc.) useful in studying pupil needs	56	7½	54	8	59	7	55	8
Av. no. of double checks	1.44		1.46		1.45		1.45	
Av. no. of single checks	3.19		3.30		2.92		3.23	

* Chi Square test of distribution of checks and double-checks on topic of fears and anxieties shows difference among groups significant at .01 level. No other significant differences.

On the three items related to possible emotional difficulty in inter-personal relations only one ("value of psychotherapy") is in the direction of our hypothesis, and the one difference significant [illegible] the .01 level of confidence shows the group-prone personalities giving higher rating to study of "fears and anxieties; how they arise and are removed". The results suggest that, quite contrary to our original hypothesis, concern over inferiority, inadequacy in personal relations, fears and anxieties may lead to greater satisfaction in the experience of being an accepted group member.

The findings on "manipulative approach" do not reveal any significant difference. "Testing" is given a low rating by both 'High's' and Low's' in attitude toward group experience; 'Discipline" rates high for both groups and what slight difference there is runs counter to our hypothesis, but is not statistically significant. The difference on "Guidance and Counseling" is even smaller. None of our hypotheses concerning personality factors in preference for group experience is substantiated by these data on choice of topics.

Further exploration of personality and background possible on the basis of a pre-test of thirty questions given to all students at the opening of the course. The first six of the thirty questions were chosen as 'best items' from the Levinson-Sanford F Scale, measuring what has come to be known as the 'Authoritarian Personality'. Hypotheses 13–17 can be checked against corresponding answers on this Pre-Test.

Hypothesis 13. High scores on the F scale items will indicate students who do not easily adapt to coöperative, democratic relationships and hence will be associated with dissatisfaction with group experience.

The results on the brief F scale, as shown in Table IX do not accord with Hypothesis 13. The association (.02 level of confidence) is in the opposite direction. The few strongly authoritarian scores—there were only sixteen cases— concentrated largely in the 'Moderately well satisfied' category. The low authoritarians (zero score) were distinctly more apt to be found in the group least well satisfied with their group experience!

Other items on the pre-test were grouped by advance inspection to permit the testing of hypotheses related to dependence of effective group work on sensitivity to emotional needs of others; absence of hostility; low self-reliance; and low 'intellectualism'.

Hypothesis 14. Seven pre-test items apparently indicative of sensitivity toward and concern over the emotional needs of others will identify students who rate group work higher.

Hypothesis 15. Four pre-test items apparently indicative of hostility and distrust toward others will identify students who rate group work lower.

Hypothesis 16. Two pre-test items apparently indicative of individualism and self-reliance will identify students who rate group work lower.

Hypothesis 17. Five pre-test items apparently indicative of 'intellectualism' will identify students who rate group work lower.

The items pertinent to Hypothesis 14 were:

a) "Feeding a child whenever he is hungry contributes to a sense of confidence in the world."

Agree: seventy-three per cent of 'High Value'; sixty per cent of 'Low Value' (Distributions not significantly different)

TABLE IX. 'AUTHORITARIAN CHARACTER' AND VALUE RATING GIVEN GROUP WORK

'Authoritarian Score'		Value Rating for Group Work			
		High	Moderate	Low	Total
High authoritarian	(4–5)*	0	14	2	16
	3*	5	21	3	29
	2	16	47	6	69
	1	21	66	18	105
Non-authoritarian	0	11	37	23	71
		53	185	52	290

$X^2 = 15.44$ significant at .02 level.

* For Chi Square computation, scores of 3, 4 and 5 were combined in a single category and Yates' correction applied to the smallest theoretical frequency. There were no scores of 6—the maximum.

b) "Behavior may be affected significantly by what a person believes to have happened even though the event never took place."

Almost no disagreement.

c) "During the first year or two of life a child is called upon to surrender psychological autonomy in favor of culturally prescribed patterns."

Agree: fifty-four per cent of 'High Value'; sixty-five per cent of 'Low Value' (Distributions not significantly different)

d) "Each child acquires a unique version of the broader culture; his private world is idiomatic."

Agree: fifty per cent of 'High Value'; sixty-three per cent of 'Low Value' (Distributions not significantly different)

e) "An educator who gave a present instead of punishment to a child who had been stealing, probably made that child's delinquent behavior more likely in the future."

Disagree: sixty per cent of 'High Value'; fifty-six per cent of 'Low Value' (Distributions not significantly different)

f) "Adolescent attitudes are more commonly influenced by the peer group than by the parent."

Agree: seventy-seven per cent of 'High Value'; eighty-three per cent of 'Low Value' (Distributions not significantly different)

g) "Aggression almost invariably arises as the result of some frustration."

Agree: eighty-one per cent of 'High Value'; seventy-one per cent of 'Low Value' (Distributions not significantly different)

Combining the seven items which appear to involve some sympathetic awareness of the emotional needs of others, we obtain practically identical scores for those who rated their group experience 'High' and those who found it of much less value. Hypothesis 14 is clearly not supported by these data.

To test Hypothesis 15, we inspect responses to the following items.

a) "The world is a hazardous place in which men are basically evil and dangerous."

Agree: two per cent of 'High Value'; two per cent of 'Low Value' (Distributions not significantly different)

b) "Young people today are much 'wilder' than they used to be."

Agree: thirteen per cent of 'High Value'; four per cent of 'Low Value' (Distributions not significantly different)

c) "Most pupils prefer to be lazy and will exert themselves on difficult tasks only under adult encouragement or pressure."

Agree: nine per cent of 'High Value'; thirteen per cent of 'Low Value' (Distributions not significantly different)

d) "Human nature the world over exhibits very much the same competitive attitudes regardless of differences in social ideals or customs."

Agree: fifty-four per cent of 'High Value'; fifty-seven per cent of 'Low Value' (Distributions not significantly different)

Again, no significant differences appear. The 'Low Value' students in this class seem not to be of the type which projects hostility onto 'people in general'.

How now about self-reliance? We have already reported that those who usually prefer to work alone rather than as a group member do not give lower ratings to their group experience. Two additional items bear on the same hypothesis.

a) "The self-reliant, completely independent person should be our educational objective."

Agree: fifteen per cent of 'High Value'; eight per cent of 'Low Value' (Distributions not significantly different)

b) "Heredity determines within narrow limits the pattern of behavio[illegible] possible to any individual."

Agree: thirty-seven per cent of 'High Value'; thirty-eight per cent of 'Low Value' (Distributions not significantly different)

No support for Hypothesis 16 can be found in these answers.

The last of the hypotheses on personality (※17) dealt with 'Intellectualism' in education as associated with rejection of learning as a group member. The following responses are pertinent.

a) "The ideal aim of education is to enable the individual to subordinate his emotion to his intellect."

Agree: twenty-five per cent of 'High Value'; thirteen per cent of 'Low Value' (Distributions not significantly different)

b) "While nursery schools and kindergartens may be attractive to children or to parents, very little of educational importance can be achieved before the child is six."

None agree.

c) "Many social attitudes depend much less upon the schooling which one has had than upon his membership within the class structure of his community."

Disagree: four per cent of 'High Value'; fourteen per cent of 'Low Value' (Distributions not significantly different)

d) "Verbal skills which tend to be significantly related to success in schools are very slightly related to success in non-school activities."

Disagree: fifty-eight per cent of 'High Value'; sixty-seven per cent of 'Low Value' (Distributions not significantly different)

e) "Persons showing high agreement with staff responses to these items will not be identical with persons who can deal most effectively with practical educational problems."

Disagree: thirty-four per cent of 'High Value'; thirty-two per cent of 'Low Value' (Distributions not significantly different)

Hypothesis 17 on 'intellectualism' is not borne out by these items; there are no differences large enough for confident prediction and on two of the five statements there is slightly more 'intellectualism' apparent in 'High' group than in the 'Low'.

On the basis of the 'Individualism' and the 'Intellectualism' hypotheses, we might expect the students who do not care for group work to give noticeably higher ratings to lectures, individual reading assignments and to writing their own professional diary. Returning for a moment to Table II, where ranks on these activities were reported, we discover that, after we have corrected for the effect upon other ranks of the high and low estimate given to 'Group Discussions', the other course activities stand in about the same order of preference. Largest differences are a liking for films by the 'Low Value' group, and approval of the professional diary by the 'High Value' group. Neither of these deviations would support the hypotheses that the group-resistant students were more bookish or more interested in independent work.

Consistently we have been disappointed in our quest. Specific answers on the pre-test do not provide the least support for apparently plausible hypotheses concerning the persons who will profit most from small group participation. A possible explanation is that the items chosen are too few and too slightly related to the trait name assigned them. It is clearly not possible to obtain a highly reliable and valid measure of personality traits when using only three or four self-report items, none of which has been validated. Perhaps the test as a whole will be more indicative than its parts. We hypothesize:

Hypothesis 18. Students with a superior background in modern psychology and education will be better able to profit from small group discussions in a course in educational psychology.—Two kinds of evidence bear on this hypothesis. One is the score on the pre-test; the other the total credit-hours from previous courses in psychology and education.

Table X indicates no significant difference between the 'High Value' and 'Low Value' individuals on their Pre-Test scores. Amount of previous study in courses in psychology and education ranged from none to over sixty credit hours. Table XI indicates that this factor also was unrelated to evaluation of group experience.

The next group of hypotheses concern differences in the particular aspects of group experience found especially satisfying or frustrating. All students ranked in order of importance five factors contributing to the success of their group and ten factors which might have been detrimental. Table XII shows that, for the class as a whole, the most valuable thing about their group experience was 'Stimulation of ideas coming from others', while 'Chance for self-expression' was rated least important to them.

What differences might be expected between 'High's' and 'Low's'? Since all items were ranked by all students the results will not reflect general level of satisfaction but only relative differences among the five satisfying aspects. It might be hypothesized that when a group goes well, little attention is given to technique. When difficulties arise, process is more closely analyzed. Hence:

Hypothesis 19. Students rating their group 'Low' in value will *attribute relatively more of such value as it has, to "learning about group process".*

A correlative hypothesis would be that a by-product of good group discussion is a warm sense of belonging and togetherness. Hence it might be anticipated that:

TABLE X. PRE-TEST SCORES AND VALUE RATING GIVEN GROUP WORK

Pre-test Scores	Value Rating for Group Work			
	High Value	Moderate Value	Low Value	Total
26–30	4	12	10	26
21–25	27	65	24	116
16–20	19	79	14	112
15 or less	3	29	4	36
	53	185	52	290

(Distributions not significantly different.)

TABLE XI. PREVIOUS STUDY OF PSYCHOLOGY AND EDUCATION IN RELATION TO VALUE RATING GIVEN GROUP WORK

Credit Hours in Psychology and Education	Value Rating for Group Work			
	High	Moderate	Low	All
0–6	4	20	3	27
7–20	14	62	12	88
21 and over	36	156	34	226
	54	238	49	341

(Distributions not significantly different.)

Hypothesis 20. Students rating their group 'High' in value will attribute relatively more of its worth to "enjoying group fellowship."

Examination of the data in Table XII does not bring much support for either hypothesis. On "learning about group process" the 'Low's' do provide sixteen 'First' rankings to only seven 'First's' from the 'High's', a difference 2.3 times its standard error. 'Second place' rankings, however, reverse the direction; and for the distribution as a whole p falls between .20 and .10.

On "enjoying group fellowship" the distributions do not differ significantly.

The one difference which does emerge clearly, in Table XII, concerns "Stimulation of ideas coming from others". This is in first place for the total group but ranks fourth among the five factors

TABLE XII. FACTORS CONTRIBUTING MOST TO VALUE OF GROUP EXPERIENCE

Stimulation of ideas coming from others	Value Rating for Group Work			
	High	Moderate	Low	All
Rank Order				
1	16	61	7	84
2	20	74	10	104
3	7	32	13	72
4	8	28	9	45
5	4	24	11	39
	55	239	50	344

(Differences significant at better than .01 level.)

Getting to know people of different backgrounds				
Rank Order				
1	9	63	13	85
2	12	49	17	78
3	17	39	10	66
4	10	54	10	74
5	7	34	1	42
	5	239	51	345

(Distributions not significantly different.)

Enjoying group fellowship				
Rank Order				
1	16	41	10	67
2	9	54	9	72
3	12	75	12	99
4	14	44	13	71
5	4	23	7	34
	55	237	51	343

(Distributions not significantly different.)

Learning about group procedure				
Rank Order				
1	7	44	16	67
2	11	37	7	55
3	8	34	7	49
4	9	57	10	76
5	9	67	11	97
	54	239	51	344

(Differences not significant: $p > .10$ and $< .20$.)

Chance for more self-expression				
Rank Order				
1	8	27	4	39
2	3	23	8	34
3	11	39	9	59
4	12	54	9	75
5	20	95	20	135
	54	238	50	342

(Distributions not significantly different.)

according to ratings from the 'Low' group. One thing most of them meant by their 'low' rating was that they didn't get much intellectual stimulation from their fellows. This finding supports the validity of the value rating and suggests a possible reason why our earlier hypotheses concerned with emotional readiness for group fellowship were not sustained. The instructors presented the group as offering opportunities for satisfying experiences in interpersonal relations; the students however persevered in the traditional set of seeking useful information from their classmates. Finding that the other group members did not contribute much in the line of intellectual stimulation, some students apparently regarded their group experience as disappointing. It should be remembered, however, that those who, on the basis of low pre-test score and few previous courses, might have been expected to have most to learn from their fellows were not significantly more appreciative.

Turning now to rankings accorded factors which hindered the group work, the over-all results (Table XIII) show the students

TABLE XIII. FACTORS DETRIMENTAL TO GROUP EXPERIENCE

Detrimental factors	Value Rating for Group Work			
	High	Moderate	Low	All
1. *Inexperience in group process*				
Rank: 1 or 2	14	85	22	121
3 or 4	18	69	15	102
5 or 6	8	37	8	53
7 or 8	7	31	4	42
9 or 10	5	15	2	22
	52	237	51	340
(Distributions not significantly different.)				
2. *Lacked solid content*				
Rank: 1 or 2	7	67	17	91
3 or 4	8	59	17	84
5 or 6	21	56	9	86
7 or 8	8	37	5	50
9 or 10	7	18	3	28
	51	237	51	339
(Distributions different at better than .01 level of significance.)				
3. *Poor physical setting*				
Rank: 1 or 2	14	64	6	84
3 or 4	15	37	8	60
5 or 6	5	43	10	58
7 or 8	11	49	15	75
9 or 10	6	42	10	58
	51	235	49	335
(Distributions different with $p > .01$ and $< .02$.)				
4. *Inadequate leaders in group*				
Rank: 1–2	5	44	20	69
3–4	14	50	13	77
5–6	8	58	12	78
7–8	14	50	5	69
9–10	10	34	1	45
	51	236	51	338
(Distributions different at better than .01 level of significance.)				

TABLE XIII—*Cont.*

Detrimental factors	Value Rating for Group Work			
	High	Moderate	Low	All
5. *Program too crowded*				
Rank: 1–2	16	52	6	74
3–4	10	55	5	70
5–6	13	61	12	86
7–8	10	43	20	73
9–10	3	25	6	34
	52	236	49	337
(Distributions different at better than .01 level of significance.)				
6. *Too little talk by some*				
Rank: 1–2	9	40	8	57
3–4	16	69	11	96
5–6	8	67	17	92
7–8	13	46	11	70
9–10	5	14	3	22
	51	236	50	337
(Distributions not significantly different.)				
7. *Too much talk by some*				
Rank: 1–2	10	41	10	61
3–4	5	53	14	72
5–6	13	61	16	90
7–8	17	55	9	81
9–10	6	27	2	35
	51	237	51	339
(Distributions not significantly different.)				
8. *Meeting time too short*				
Rank: 1–2	22	40	0	62
3–4	9	32	6	47
5–6	11	28	4	43
7–8	5	48	13	66
9–10	3	86	26	115
	50	234	49	333
(Distributions different at better than .01 level of significance.)				

TABLE XIII—*Cont.*

Detrimental factors	Value Rating for Group Work			
	High	Moderate	Low	All
9. *Too little staff direction*				
Rank: 1–2	4	71	11	86
3–4	7	64	10	81
5–6	8	47	10	65
7–8	11	40	11	62
9–10	20	11	7	38
	50	233	49	332

(Distributions different at better than .01 level of significance.)

Detrimental factors	High	Moderate	Low	All
10. *Too much staff direction*				
Rank: 1–2	5	10	2	17
3–4	1	8	2	11
5–6	5	12	3	20
7–8	4	35	6	45
9–10	35	167	36	238
	50	232	49	331

(Distributions not significantly different.)

inclined to place most blame on their own "inexperience in group process".

The second criticism in order of importance was "not enough solid content in the discussions". Students on the whole did not feel that there had been too much or too little staff direction. Although training for group discussion often emphasizes the handicap imposed by the member who talks too much or too little, these were not felt by this class to be major liabilities.

Differential rankings of the ten potential liabilities to good group functioning may be analyzed with reference to a hypothesis that those well satisfied will play down the vital criticisms and play up the more external and less significant liabilities, while those dissatisfied will regard the vital and significant factors as major liabilities. More specifically:

Hypothesis 21. Members who rate group value as 'Low' will give relatively more emphasis to charges that the discussion lacked solid content and good leadership; those who rate group value as 'High' *will give top rank to such 'external' criticisms as "Meeting time too short", "Physical setting poor", and "Schedule too crowded".*

The data of Table XIII fully support this hypothesis. There were statistically significant differences on all of the items mentioned, and consistently in the predicted direction.

A few more hypotheses may be derived from consideration of certain personal data. For instance, the anthropologist Ashley Montagu in recent articles has argued that women are basically less aggressive and more coöperatively disposed than are men. So we might predict that:

Hypothesis 22. Women will value their group experience more highly than will men.

TABLE XIV. SEX DIFFERENCE AND VALUE RATING GIVEN GROUP WORK

	Value Rating for Group Work			
	High	Moderate	Low	Total
Men	42	132	22	196
Women	13	111	29	153
	55	243	51	349

(Distributions different at .01 level.)

The data on Table XIV reveal a statistically significant difference, but the difference shows a higher proportion of men in the category of 'High approval'. Our sex-difference hypothesis must be reversed. Why were the men better satisfied? One possibility would be that when groups are formed, as these were, by a process which leaves it to students somehow to group themselves, the men take more initiative and are less apt to be drawn in with unwanted companions. Another factor may be the greater deference shown to the males by members of a mixed group in our culture. Still another explanation may be that, contrary to the stereotypes of humor, men in such groups talk more than women do. Our data do not permit us to check the plausible hypothesis that level of satisfaction reflects the level of participation.

There remains the variable of age and experience. Some students were fresh out of college; others had taught for more than twenty years. Should we expect a difference in attitude toward group discussion among students at widely different stages of professional maturity? We argued that students just beginning their teaching would probably feel they were getting more out of conferring with experienced colleagues than the older teachers would think they could get from youth. Moreover, if small group work may be considered still to be an innovation, it would be more readily accepted by the younger students. The pattern of respectful listening to professional authority may be more firmly established in older teachers. Hence we predict that:

Hypothesis 23. Students with no professional experience are more likely to value group work highly; those with long practical experience are more likely to reject or to disparage the small group discussions.

TABLE XV. YEARS OF TEACHING EXPERIENCE AND VALUE RATING GIVEN GROUP WORK

Years of Experience	Value Rating for Group Work			
	High	Moderate	Low	Total
None	22	64	6	92
1–5 years	23	93	21	137
6–10 years	2	45	15	62
11–20 years*	6*	28*	8*	42*
More than 20 years*	0*	10*	1*	11*
	53	240	51	344

(Distributions significantly different at .01 level.)
* Combined for Chi Square test.

The data of Table XV are in accord with the hypothesis. As many as forty-two per cent of the 'High Value' group were without any previous teaching experience, as compared with only twelve per cent of the 'Low Value' participants. The same conclusion is supported also by other break-downs showing association between 'Low Value' ratings and having present full-time employment, mainly as a teacher ($p > .02$ and $< .05$), and also between 'Low Value' ratings and having as professional goal administration or supervision rather than teaching ($p < .01$). The group work was apparently most helpful to full-time, younger students, inexperienced in teaching, and eager to profit from the broader experience of others. The experience factor, for reasons not clear, was oddly sex-linked. High ratings came predominantly from inexperienced men; low ratings from women with more than five years teaching experience. Proportion of high-low was about 50–50 among experienced men or among inexperienced women.

SUMMARY AND INTERPRETATION

Probably the most striking revelation of this report is the extent of mistaken anticipations in the writer after a quarter century of organizing, training and observing small group discussions as part of large graduate courses. Of our twenty-three logically plausible hypotheses only two were confirmed; seventeen collapsed for want of factual support, and four had to be reversed in whole or in part.

Students who will enjoy and profit from small group participation could not be identified on the basis of; their own expressed preference; their level of mastery of the course material; their stated interest to learn about 'group leadership'; their general level of enthusiasm for course topics; or their responses to clusters of questions apparently indicating sympathy, hostility, self-reliance, or 'intellectualism'. No advantage was demonstrated for groups which worked cumulatively on a single topic all semester over groups which discussed different issues each week.

Obtained differences which were statistically significant demonstrated that:

1) individual-wide variables accounted for value ratings better than did group-wide variables. Few groups were homogeneously regarded as good or poor;

2) the best groups were larger (ten–fifteen members) than average (eight);

3) Students especially interested to learn about 'fears and anxieties' tended to place high value on group work;

4) Students who rejected all items indicative of 'authoritarianism' placed a low value on group work;

5) Students who rated groups low were disappointed mainly in lack of intellectual stimulation from their fellow-members;

6) Men, with little or no professional experience in the field, were responsible for more 'High' ratings; women with more than five years of experience gave more 'Low' ratings.

ELEVEN-YEAR-OLD BOYS IN TROUBLE[1]

WILLIAM W. WATTENBERG

Wayne University

Among workers in the field of juvenile delinquency, we find much attention given to three common but not mutually exclusive explanations of why youngsters fall into patterns of misconduct. Sociologists have tended to concentrate upon factors linked to social or economic variables. Such workers as Shaw and his collaborators (*5*) have painstakingly produced evidence showing that delinquency rates are highest in neighborhoods typified by poverty, low status, poor housing, and cultural conflict. According to their view, delinquency is a relatively normal reaction of young folks to bad situations.

A second line of evidence receives more attention from psychologists and psychiatrists. They have found that among delinquents there are many who prove to be victims of emotional conflict or instability. Healy and Bonner (*3*) have shown that where delinquent youngsters had non-delinquent siblings, presumably exposed to the same social environment, the delinquents differed from their siblings in being emotionally maladjusted. More recently, Redl and Wineman (*4*) have pictured in great detail the personality disorganization in a group of very seriously delinquent boys.

A third type of causal situation, less fully explored by research workers, has been stated by Washburne (*6*). He sees delinquency as often a product of a child's inability to apply judgment to the control of his impulses. As this could be a temporary condition, which would change itself as normal maturing brought increased power of judgment, it can account handily for the fact that many boys and girls engage in serious misconduct for a while, later correct their anti-social tendencies without benefit of psychotherapy or neighborhood reform.

---

[1] The author wishes to express his indebtedness to Senior Inspector Sanford Shoults, Inspector Ralph Baker and Lieutenant Francis Davey of the Youth Bureau, Detroit Police Department, for their coöperation in making available the records upon which this study is based. Appreciation is also due to Dean Waldo Lessenger, of the College of Education, Wayne University, for solving administrative problems permitting completion of the research.

Clearly, these three explanations of delinquency need not be mutually exclusive. Each could be operating either alone or in combination for some cases. In the total mass of juvenile delinquents we could reasonably expect to find some with seriously disturbed personalities, some normal youngsters who had learned patterns of misconduct by living in unfortunate social settings, and some who had committed misdeeds because unaware of the full implications of their acts.

The focus of the present study is upon very young boys in trouble. In terms of the now popular terminology of developmental phases, they are preadolescents. This group was chosen because it has received relatively little systematic study and because its characteristics could be expected to throw additional light on the genesis of delinquency.

In their summary of research on preadolescents, Blair and Burton (*1*) point out a number of qualities which might lead to delinquency taking somewhat different patterns from that found in older age groups. In terms of attitudes toward adults, the preadolescents might be more likely to display ambivalence. Among boys, the search for masculine identification objects gives greater power to gang codes. They are given to seeking models for behavior among the next older age group. Their immaturity at a time when impulses are rising in power is likely to lead to periods when their value systems are somewhat hazy and weak.

Related to the three lines of explanation previously mentioned, we would predict the following factors as likely to be found among preadolescents involved in serious misconduct: Because of their tendency to ape older adolescents, we would expect that bad neighborhood conditions, in which delinquency was part of the subculture, would be very influential. Because of the temporary weakness of their value systems, we would expect to find in the total group a somewhat higher proportion of relatively normal youngsters than among an older group of delinquents. We should also expect to find some young folks giving evidence of serious maladjustment, although they would be present in less heavy a concentration.

In a study of factors related to repeating, reported elsewhere

(*8*), the author had noted that among eleven-year-olds with police records repeating was more highly predicted by school grades and by the nature of gang activities, and less highly predicted by family conditions and social variables than among delinquents unselected as to age. This finding, however, did not give an answer as to the relative frequency in the total group of youngsters upon whom the several influences associated with delinquency might be operating.

PROCEDURE

It was decided to compare a group of eleven-year-olds known to police, as the result of complaints, with a similar group who had passed their twelfth birthdays. The age eleven was chosen because it was believed a sufficiently large number of cases would be found to permit statistically reliable computations and because the results of research such as the Harvard Growth Study (*2*) indicated that few eleven-year-old boys have advanced far in pubescence. In the sample of seven hundred and forty-seven boys none had completed his preadolescent growth spurt before 11.50 years of age, and only thirteen in the 11.50–12.49 annual period.

Although the eleven-year-olds could not be considered a pure sample of preadolescents nor the older group, of adolescents, yet we can say safely that the younger sample would contain a very heavy concentration of preadolescents and the older sample would be much more strongly saturated with adolescents.

The files for 1949 of the Youth Bureau of the Detroit Police Department were searched and the records of all boys more than eleven and less than seventeen years old obtained. There was a total of 4,121 such records. On the basis of age of boy at the time of police contact these were divided into two groups, consisting of 334 eleven-year-olds and 3,787 boys past their twelfth birthdays.

For every boy there was already in the file a 'history sheet', filled out by specially assigned officers on the basis of interviews with the boys, visits to his home, and their own knowledge of conditions in the districts which they regularly covered. These sheets contained some forty-two items of fact or rating concerning the boy, his housing, his school, his family, and his neighborhood. For every item, a tabulation was prepared and the chi-square calculated.

The data on the history sheets have been used in a number of previous studies (*7*, *9*) and found to be reliably predictive of repeating and to yield valid discriminations between gang members and non-members.

FINDINGS

The extent of differences between the eleven-year-olds and the older boys is indicated by the fact that of the forty-two chi-square calculations, in nineteen the null hypothesis could be rejected at the two per cent level of confidence. By chance alone, one such rejection could be expected to occur. Thus, the number of significant tables was eighteen times above chance expectation. We can therefore state with assurance that the eleven-year-olds as a group showed some marked contrasts to the older group. For convenience, these differences will be reported in clusters having elements in common.

Socio-economic indices.—Of the data bearing on socio-economic factors or neighborhood conditions, the most objective was the ratio of rooms in the dwelling unit to number of occupants. The facts are presented in the upper third of Table I. This shows a statistically reliable difference to exist. (In this table, if the 'not stated' category is eliminated, the chi-square total is 7.6; with one degree of freedom, P is less than 0.01.) The eleven-year-olds came in larger proportion from dwellings with one room or fewer per occupant. Bearing out this table were a number of others. The police officers found more eleven-year-olds in buildings rated substandard; on the chi-square test this relationship was significant at a one per cent level of confidence. A similar trend was also noted in the case of type of building, mixture of business and residential land usage, and rated quality of neighborhood, although in these instances the chi-square test was inconclusive. The younger group came in greater proportion from rented quarters, neighborhoods rated 'below average', and areas where business establishments were mixed with residences.

Family conditions.—As far as having both parents present in the home, our younger group proved to have the advantage. The middle third of Table I gives data showing that more of them came from intact homes; more of the older group had lost one or more parents by death. The relationship was strong enough to warrant rejection of the null hypothesis at the 0.001 level of confidence. This relationship was supported by other tables. At the two per cent level of confidence, more of the older group had working mothers, as contrasted to the eleven-year-olds in whose families the father was more likely to be the only employed parent. Among the tables where the chi-square test was inconclusive, the eleven-year-olds more often came from homes where some parent was at home to take care of things during the day and evening, as contrasted to homes unsupervised except at night. On the basis of interviews with the boy and his parents the officers' report showed a statistically inconclusive tendency for the older group to have parents who frequently quarreled.

TABLE I.—HOME CONDITIONS OF BOYS

	Total	11-year-olds	Older group	χ^2	df	*P*
Ratio of rooms to occupants in dwelling units						
One room or less per person	3,127	276	2,851			
More than one room per person	931	55	876	9.0	2	<0.02
Not stated	63	3	60			
Marital status of parents						
Living together	2,456	223	2,233			
Separated or divorced	971	81	890	17.6	3	<0.001
One or both dead	645	26	619			
Not stated	49	4	45			
Method used by parents in giving money to boy						
Allowance	931	93	838			
Pay for work	98	2	96	64.3	4	<0.001
On request	2,152	218	1,934			
None	809	18	791			
Not stated	131	3	128			
Total	4,121	334	3,787			

Dependency of boys.—The only item giving a clue to dependency related to the manner in which the boys received their spending money. The bottom third of Table I sets forth the data on this point. Here it will be noted that the younger boys, with few exceptions, received money from parents either as an allowance or in response to their requests. By contrast, the older boys included a sizable minority whose parents gave them no money or else paid a wage for work done. This relationship was significant at the 0.001 level of confidence whether or not the 'not stated' category was treated as a separate class. The 'not stated' group is reported, as it will be discussed later.

Expressed attitudes toward institutions.—When interviewing the boys, the police officers sought information indicating their attitudes toward home, school, church, and adult neighbors. Obviously, what the boy would say to a policeman on items where 'good' and 'bad' were so patent must be taken with a grain of salt. As a correction to falsifications, efforts were made to secure objective behavior. In general, on both types of evidence the eleven-year-olds more often gave the conventionally acceptable response or showed more conventional behavior. In the upper half of Table 2 we give a mixture of both types of data relating to school. Here we note that the older group was more likely to openly express dislike of school. In addition, a significant minority had quit formal education. Other tabulations gave results showing a similar trend. At the one per cent level of confidence, more of the older group expressed hostile feelings toward teachers and one or both parents. Among the statistically inconclusive tabulations, we found more older boys antagonistic to adult neighbors and more eleven-year-olds attending church regularly.

Peer group relationships.—The younger boys appeared to have, as a group, better social relationships with other youngsters. Among the facts to which the police officers gave close attention was the boy's companions. Data on this point were valued highly for police reasons; they had often proved useful in clearing up new offenses. The facts were obtained not only from the boy himself but from patrolmen and other adults familiar with the neighborhood. On this basis it was possible to divide the boys into two groups: (1) those who belonged to a crowd or a gang which played together and (2) 'lone wolves'. The lower half of Table 2 presents the data on this point. The eleven-year-olds included more boys who belonged to a crowd or gang and fewer who had been classed as 'lone wolves'. As with the other key tabulations reported above, this one also was supported by other data. At the one per cent level of confidence, the eleven-year-olds were reported to get along better and to be less given to quarrels with classmates in school.

Also, in their choice of favorite sports, comparatively more chose baseball and fewer, fishing, than the older group.

Miscellaneous.—There were a number of other items yielding statistically significant differences between the two age groups. For the most part these either were not germane to the main purpose of this study or else were natural results of age differences. For the sake of completeness, they are reported herewith in summary form:

1) Fewer eleven-year-olds had paid employment.

TABLE 2.—SCHOOL ATTITUDES AND PEER GROUP RELATIONSHIPS OF BOYS

	Total	11-year-olds	Older group	χ^2	df	P
Attitude expressed toward school						
Likes	2,675	271	2,404	58.1	4	<0.001
Indifferent	591	42	549			
Dislikes	369	17	352			
Hates	60	2	58			
Not stated and not in school	426	2	424			
Peer group membership						
Boy was in some sort of peer group	3,632	313	3,319	11.4	2	<0.01
'Lone wolves'	473	20	453			
Not stated	16	1	15			
Total	4,121	334	3,787			

2) Police officers rated more of them small for their age.

3) Police officers considered more of them 'honest'.

4) On the basis of appearance, police officers rated more of them 'preadolescent'.

5) More eleven-year-olds spent all their money on entertainment.

6) None was allowed to drive the family car.

DISCUSSION

On one of the hypotheses, the results were unequivocal. As indicated by the housing situation and neighborhood ratings, low socio-economic status was found in a higher proportion of the eleven-year-olds.

As to the relative proportion of seriously maladjusted youngsters among the eleven-year-olds, the evidence is indirect. A previous study (*9*) in this series had revealed that among repeaters giving evidence of personality distortion, strongly concentrated among boys who were in trouble over and over again even though not members of gangs, certain items were linked with recidivism and reliably predictive of it. Among these were the boys' expressed attitude toward parents, the presence of family tension as evidenced by broken homes, and the failure to supply information on such practices as the giving of money by parents. The last-mentioned was striking. It could be interpreted either as an emotional blocking, as an evidence of shame, or a desire to be close-mouthed about family affairs. In any event, in the present study, all these indices previously found associated with maladjustment were less frequent in the eleven-year-old segment of the sample. Although the evidence leans in the direction predicted by our hypotheses, it obviously requires verification by studies utilizing more direct and sensitive measures of emotional stability.

One result of the present study was not fully anticipated. This had to do with the greater conventionality in the eleven-year-olds' expressed attitudes toward parents and teachers. If this group had hostile feelings, as we would expect they should, these were but one side of an ambivalence. Perhaps the fact that they were still more dependent, as indicated by their having to rely upon parents for spending money, made them more apt to feel uneasy at the prospect of open defiance. At this point we can only speculate. Several possibilities suggest themselves. It may be that fewer of these youngsters are emotionally disturbed in any marked degree and that their conventionality is merely a sign of that fact. It is equally possible that the conventionality is merely a residue of childish attitudes which they will out-grow, and, as years bring them added independence, rebellion will become more wholehearted. Here, again, there is need for further investigation in which such sensitive devices as various projective tests can give us a clearer picture of what lies below the surface.

SUMMARY

The police files on all boys interviewed on complaint by Detroit Youth Bureau officers in 1949 were studied, and the 334 eleven-year-olds compared with the 3,787 who had passed their twelfth

birthdays. The eleven-year-olds as a group were found to come in greater proportion from poorer socio-economic levels, to be more dependent upon the parents, to express more conventional attitudes toward adult-managed institutions, and to have better social relationships with other youngsters.

REFERENCES

1) Arthur Witt Blair and William H. Burton. *Growth and Development of the Preadolescent*. New York: Appleton-Century-Crofts, 1951.

2) Walter F. Dearborn, John W. M. Rothney, and Frank W. Shuttleworth. "Data on the Growth of Public School Children (From Materials of the Harvard Growth Study)," *Monographs of the Society for Research in Child Development*, Vol. 3, No. 1, 1938.

3) William Healy and Augusta F. Bronner. *New Light on Delinquency and Its Treatment*. New Haven: Yale University Press, 1936.

4) Fritz Redl and David Wineman. *Children Who Hate*. Glencoe: Free Press, 1951.

5) Clifford R. Shaw and Henry D. McKay. *Juvenile Delinquency and Urban Areas*. Chicago: University of Chicago Press, 1942.

6) J. N. Washburne. "Definitions in character measurements." *Journal of Social Psychology*, Vol. 1, pp. 114–119, 1931.

7) William W. Wattenberg. "A Comparison of Repeaters and Non-repeaters Among Boys in Trouble with the Police in Detroit in 1946 and 1947," *Papers of the Michigan Academy of Science, Arts, and Letters*, Vol. 35, pp. 395–405, 1949.

8) ——— "Factors associated with repeating among preadolescent 'delinquents'." (Accepted for publication by the *Journal of Genetic Psychology*.)

9) William W. Wattenberg and James J. Balistrieri. "Gang membership and juvenile misconduct." *American Sociological Review*, Vol. 15, pp. 744–752, 1950.

THE RELATIVE INTELLIGIBILITY OF MALE AND FEMALE TALKERS[1]

B. SILVERSTEIN, R. C. BILGER, T. D. HANLEY, AND M. D. STEER

Purdue University

Very early in the history of experimental investigation in speech, researchers recognized that males and females constitute different experimental samples. The two sexes cannot be assumed to be drawn from the same population when some of the voice variables (notably pitch) are considered. Accordingly, most of the research articles to be found in publications in the speech field report results of investigations employing all male experimental samples. A few experiments employing female subjects have been reported. However, experiments involving both male and female subjects are almost unprecedented.

The paucity of speech research involving female and male-and-female groups is particularly to be noted in the series of intelligibility studies conducted under military auspices during and after the Second World War. Many important results of such investigations, applicable only to a male population, have been published. Among the findings which have been demonstrated to have statistical significance are the following:

1) Speech signals louder than conversational level are necessary to intelligible communication in noise (*1*, *2*).

2) Instruction and practice in increased syllable duration will result in improved intelligibility (*7*).

3) Practice involving 'Read Back' of transmitted messages through difficult transmission conditions will result in improved intelligibility (*5*).

4) Taking intelligibility tests will improve intelligibility (*6*).

5) Two hours of instruction in loudness and clearness will result in a substantial improvement in speech intelligibility (*7*).

6) Intelligibility tests given under laboratory conditions are valid for distinguishing the superior from the inferior speakers, with respect to speech intelligibility in high level noise (*4*).

---

[1] This research was carried out under contract with the Office of Naval Research, Special Devices Center, Human Engineering Division, as Contract N6ori-104, T. O. II, Project NR-782-003, of which this is Technical Report Number SDC 104-2-29.

However, as previously stated, in all of the investigations from which these conclusions were drawn, the samples employed were limited to males of military age. Since the population concerned with voice communication in high level noise was almost exclusively male during the first period of these investigations, it was practically imperative that the samples be exclusively male. Today, however, an increasing tendency exists for the assignment of women to responsible communication positions, within the armed services and in volunteer civil defense agencies. Illustrative of this increased participation by women in the defense effort is the number of women currently engaged in control tower duties at civilian and military airfields.

The increased dependence upon women for vital military voice communications makes necessary a reëxamination of the known facts about intelligibility to determine their applicability to female talkers. It is important, therefore, to ascertain whether or not voice communication normative data drawn from investigations involving all male subjects may be applied to the female population which may be called upon to do vital communication work. Considered to be particularly pressing is the question of how well female talkers perform when in the presence of high level noise.

The purpose of this investigation was to determine the relative intelligibility of male and female talkers over standard military communication equipment in the presence of high level noise. More specifically, it was of interest to determine if there were any differences in speaking ability manifested between untrained groups of male and female talkers; if there were any differences in the effects of training for improved intelligibility upon groups of male and female talkers; and if there were any differences in intelligibility scores of male or female talkers attributable to the sex of the listening panel.

SUBJECTS

Initially the subjects for this investigation were ninety male and ninety female students enrolled in sections of an elementary course in public speaking at Purdue University. Of these one hundred and eighty subjects, forty-five male and forty-five female subjects were designated as the experimental group to be given two hours of training in voice communication in addition to being tested. The remaining subjects were designated as the control group and were tested and retested without intervening training. Further division of both the control and experimental groups was effected so that there were three subdivisions within each group: all male panels, all female panels, and panels composed of equal numbers of men and women.

Of the initial group of one hundred and eighty subjects, forty-two were dropped from the investigation because of failure to attend training or retesting sessions, or because their scores were selected at random as scores to be discarded in order to equate the sizes of the subgroups. (In order to employ a standard form of orthogonal analysis it was necessary to maintain equal-sized subgroups in the experimental and control samples.)

PROCEDURE

Instrumentation.—Three Portable Interphone Trainers (Navy Device 8-I) were utilized in this investigation. One Device 8-I was used as a noise generator and the noise from this source was fed through a junction box into the phono input of two other Device 8-I's which were used as Interphone Trainers (Fig. 1). The amplified noise from each Device 8-I being used as an interphone trainer was then coupled to ten sets of headphones, Model ANB-H-1. The noise level output was adjusted so that the 'full noise' condition produced a noise level of 106.5 db, re 10^{-16} watts/cm^2, in the headphones, as measured by an ADC Artificial Ear. This noise level, used in the intelligibility testing sessions, was set initially by three judges on a subjective criterion. It was judged to be a noise level which would allow untrained speakers to achieve approximately fifty per cent intelligibility on the VCL, 24-Work Multiple-Choice Test lists. (*5*).

The speech input from the carbon microphones, Model T-38C, was coupled to the Channel No. 1 carbon microphone input of each Device 8-I being used as an Interphone Trainer. The speech channel gain was calibrated so that an input of 0.07 volts, at 1,000 cps, produced an output of 0.49 volts across the headphones.

For the 'reduced noise' condition, used during the training session, the noise level was reduced 10 db, to 95.6 db, re 10^{-16} watts/cm^2. The speech channel gain was also reduced by 10 db in order to maintain the same signal-to-noise ratio that was used in the testing situation with 'full noise'.

Method.—The procedure followed in this study is similar to the general pattern followed in several other studies of talker intelligibility in the presence of high level noise previously reported from this laboratory. (*6*, *7*, *8*) The procedure in this investigation may be outlined as follows:

A) A pre-training test of intelligibility given to all subjects;
B) A training period for the experimental group only; and
C) A post-training test of intelligibility for all subjects.

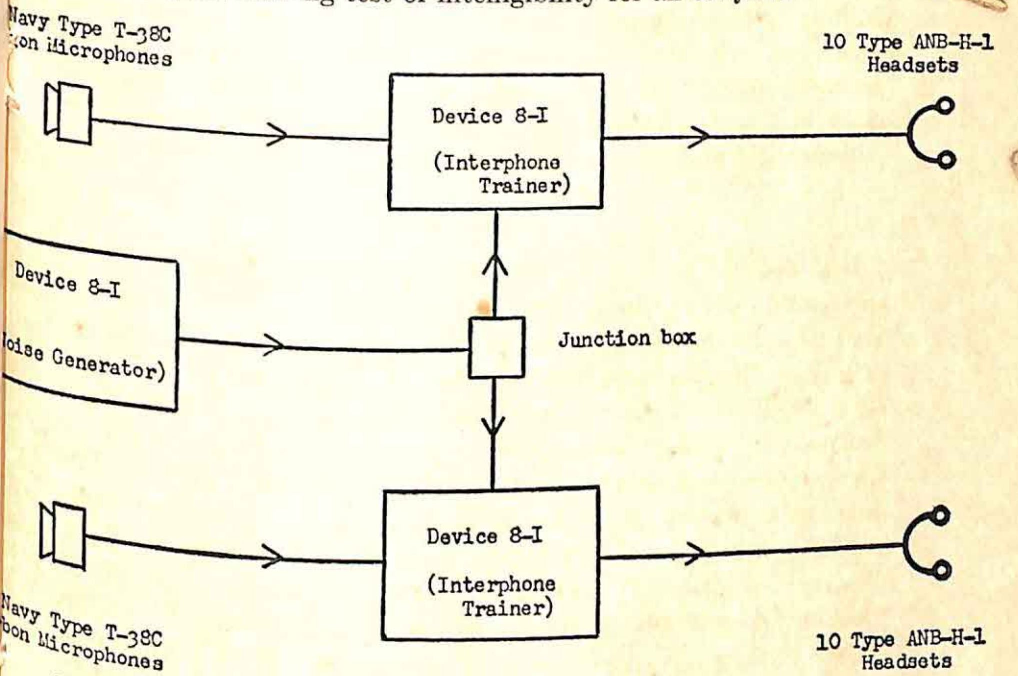


FIG. 1. Block diagram of instrumentation used to test speaker intelligibility

Pre-training test.—The subjects came to the testing room directly from their classrooms. They had been organized into specific test panels of from seven to ten members previous to their introduction into the testing situation. Each panel member was assigned a seat from among ten straightback, tablet armchairs arranged in a straight line. The seats were partially enclosed by booths of celotex sheeting mounted between the chairs. Generally, only one panel was tested at a time, but it was possible to test two panels simultaneously by using both of the interphone trainers.

After being seated, the subjects were given a brief period of orientation to the task and the equipment. The panel members were told, also, that they were participating in an investigation being carried out by Purdue University for the Office of Naval Research; that the investigation was concerned with voice communication in the presence of high level noise, and that they would serve both as speakers and listeners during the investigation. Following instruction in the proper use of the microphone and headphones, the subjects were instructed to put on their headsets and they were given specific instructions over the interphone trainer for the VCL, 24-Work Multiple-Choice Intelligibility Test. (*4*) After the instructions had been given, the 'reduced noise' condition was transmitted to the headphones and the experimenter read a sample intelligibility test. Following this, the 'full noise' condition was transmitted to the headphones, and Form A of the 24-Word Multiple-Choice Intelligibility Test was administered. The subjects were dismissed immediately after testing was completed.

Training.—Three to four weeks after the administration of the pre-training test, all of the experimental subjects were called back with their original panel mates and given a one-hour training session. The first part of the training session consisted of a lecture pointing up the importance of loudness and clear pronunciation in radio-telephone communication.[2]

The training lecture emphasized the use of a loudness level 'just short of shouting' for maximum intelligibility. It was pointed out that the speaker might use the speech signal in his own headphones (sidetone) to determine whether he was using sufficient loudness for the particular noise barrier. In addition to these instructions, the correct use of the microphone was emphasized again. Instructions for increased clearness included the suggestion that all of the sounds of the words should be uttered precisely and that all of the syllables should be given equal loudness and equal duration.

Following this brief lecture period the subjects were given an opportunity to practice the techniques which had been discussed. Each subject was given a printed military-type message which he was directed to read to another subject in his circuit.[3] The specified listener called for repeats until he was certain of the message.

---

[2] A detailed description of this lecture can be found in SDC Technical Report, No. 104-2-4, Purdue Voice Science Laboratory, Lafayette, Indiana, 1947.

[3] This departure from the procedure of previous investigations conducted at this laboratory (*6*, *7*, *8*) is to be noted. Hitherto the training messages utilized consisted of non-military communications.

Then he repeated it for the originator, who corrected any errors in message content. After the message recipient read back the message correctly, another pair of subjects was given an opportunity to engage in similar practice. During this training period each subject acted as the originator and as the recipient of a message to be 'read back' correctly. This practice period was interspersed with further instructions and specific criticisms in which was stressed the importance of the techniques being employed for increased intelligibility.

Original messages and repeat-backs were spoken over the same interphone circuits used in testing. The noise level during the training program was set on 'reduced noise' condition.

Post-training test.—Three to four weeks after the experimental group training sessions, post-training tests were administered to both the experimental and control subjects. All subjects were tested with their original panel-mates and all subjects occupied the same seats and used the same equipment that they had used in previous sessions. Form B of the VCL, 24-Word Multiple-Choice Intelligibility Test was administered to all subjects as the post-training test. All calibrations were identical to those used in the pre-training test.

Statistical analysis.—As previously indicated, the purpose of this study was to investigate certain aspects of the performance of male and female talkers, specifically their training. For statistical analysis, per cent intelligibility scores for individuals and subgroups were computed. The mean number of items correctly marked by a listening panel for a given talker, divided by the number of words spoken by the talker (twenty-four) constituted the intelligibility score for the talker.

The data collected were treated by three separate analyses. First, the data from the pre-training test were analyzed by an analysis of variance technique to determine if there were any statistically significant differences, with respect to speaker intelligibility, between the sex subgroups at the outset of the experimentation. The data from the post-training test were treated by the same analysis of variance design to determine if any differences were present in the post-training test results. Finally, so that the comparison would be based on the 'best-fitting' regression line rather than the assumption of a perfect regression, an analysis of covariance was used to compare pre-training to post-training scores in order to determine if any significant differences were present in the intelligibility gains made by the sex subgroups.

Whenever significant differences were found, 't' tests for the significance of difference between means were used to isolate the differences.

TABLE I.—SUBGROUP MEANS AND STANDARD DEVIATIONS FOR TEST AND RETEST

Subgroup	N	Test	SD	Retest	SD
Experimental Group (Training)					
Male	23	57.1	10.6	69.2	10.2
Female	23	52.3	12.6	69.0	8.3
Mixed	23	53.2	9.9	68.2	9.0
Total	69	54.2	11.1	68.5	9.2
Control Group (No-Training)					
Male	23	64.9	11.0	62.8	12.4
Female	23	48.9	9.0	53.5	8.9
Mixed	23	53.0	9.5	59.1	9.8
Total	69	55.6	9.9	58.5	10.5

RESULTS

The means and standard deviations of the intelligibility scores for all subgroups, experimental and control, on both the pre-training and post-training intelligibility tests are presented in Table I.

Relative Intelligibility of Untrained Male and Female Speakers.—The results of the analysis of variance utilized to test the significance of differences in the pre-test data are presented in Table II. The 'F-ratio' for between sex subgroups is significant at the one per cent level; but, before any conclusions may be drawn, the significant interaction, experimental-control X sex subgroups, must be noted, investigated and interpreted.

Since the interaction experimental-control X sex subgroups is significant at the 2.5 per cent level, it is permissible to conduct separate analyses of variance on each of the main groups in order to investigate the source of the interaction. The results of these analyses (see Table III) indicate that the reason for the significant interaction was that there were differences, significant beyond the one per cent level, between the sex subgroups within the control group, while there were no significant differences within the experimental group. However, the magnitude of the 'F-ratio' in the case of the control group is so great that it may be reasoned that the significant interaction is the result of chance allocation of subjects to experimental or control group.

TABLE II.—ANALYSIS OF VARIANCE TO TEST INITIAL DIFFERENCES IN INTELLIGIBILITY

Source of Variation	Sum of Squares	df	Mean Squares	'F-ratio'	Probability
Between sex subgroups	2,724.26	2	1,362.13	11.80	P < .001
Between experimental-control	67.91	1	67.91	<1	—
Interaction: experimental-control X sex subgroups	759.52	2	379.76	3.92	.01 < P < .025
Within groups (error)	15,239.00	132	115.45		

TABLE III.—ANALYSIS OF VARIANCE TO TEST INITIAL INTELLIGIBILITY DIFFERENCES BETWEEN SEX SUBGROUPS WITHIN THE EXPERIMENTAL AND CONTROL GROUPS

Group	Source of variation	Sum of squares	df	Mean square	'F-ratio'	Probability
Experimental	Between sex subgroups	305.31	2	152.46	1.29	0.25 < P < 0.50
	Within sex subgroups	7,784.19	66	117.94		
	Total	8,089.50	68			
Control	Between sex subgroups	3,178.49	2	1,589.24	15.57	P < 0.001
	Within sex subgroups	6,734.78	66	102.04		
	Total	9,913.27	68			

In summary, the results of analysis for sex differences in intelligibility prior to training reveal significant differences favoring the male sex in the control group. That a similar result was not found in the experimental group is believed to be a chance effect related to assignment of subjects to groups.

TABLE IV.—ANALYSIS OF VARIANCE TO TEST FOR DIFFERENCES IN POST-TRAINING TEST SCORES

Source of variation	Sums of squares	df	Mean squares	'F-ratio'	Probability
Between sex subgroups..	587.48	2	293.74	2.87	$0.05 < P < .10$
Between groups (experimental-control).......	3,426.04	1	3,426.04	33.44	$P < 0.001$
Interaction: experimental-control X sex subgroups................	428.19	2	214.10	2.09	$.10 < P < .25$
Within groups (error)...	13,522.39	132	102.44		
Total.............	17,964.10	137			

TABLE V.—ANALYSIS OF COVARIANCE TO TEST DIFFERENCES IN IMPROVEMENT BETWEEN INITIAL AND FINAL TEST

Source of variation	Sums of squares	df	Mean squares (adjusted scores)	'F-ratio'	Probability
Between sex subgroups..	123.10	2	68.55	—	—
Between experimental-control.............	3,762.87	1	3,762.87	41.05	$P < .001$
Interaction: experimental-control X sex subgroups...............	154.81	2	77.40	—	—
Within cells (error).....	12,008.64	131	91.67		

Effect of mixed listening panels upon speaker intelligibility, pre-training test.—In an attempt to determine the effect a mixed listening panel could have upon speaker intelligibility scores, 't' tests for the significance of differences between means were calculated as follows:

(a) All male subgroups vs. males in mixed subgroups, and

(b) All female subgroups vs. females in mixed subgroups. Neither of the 't-ratios' approached significance.

Sex subgroups speaker intelligibility differences in post-training test.—The results of the analysis of variance used to test for the significance of differences on the post-training test (see Table IV) indicate that there were no significant differences between sex subgroups on the post-training test. The highly significant difference between experimental and control groups is the result which, on the basis of previous research (*5*), was expected.

Sex subgroup speaker intelligibility gains.—Table V presents the results of the analysis of covariance used to determine if any significant differences existed in the relative degree of improvement made by the sex subgroups. The failure of the 'F-ratio' for between sex subgroups to reach a significant level indicates that, when initial differences in speaker intelligibility are taken into account, males and females benefit equally from training (whether training consists of the one-hour lesson or merely of a retest).

CONCLUSIONS

Within the limitations of this investigation, the following conclusions seem justified:

1) The existence of sex differences, with respect to the ability of untrained talkers to speak intelligibly in the presence of high level noise, is demonstrated by this investigation. Differences favor the male sex.

2) Wherever these sex differences are found, training, if only the training incidental to undergoing a retest of speaker intelligibility, serves to eliminate these differences.

3) A one-hour period of training for improved intelligibility results in a significantly greater gain in intelligibility than that made by a control group on test-retest. This conclusion is applicable to all male, all female, and mixed groups.

4) Taking a test for speaker intelligibility in the presence of high level noise results in statistically significant gains in intelligibility on subsequent tests. However, when initial scores are high, the trend is not consistent and may even be reversed.

5) The sex of the auditor seems to have little or no effect upon the intelligibility scores of male and female speakers.

6) Male and female speakers benefit equally from intelligibility training.

REFERENCES

1) Harry M. Mason, "The effect of very loud speech signals upon intelligibility," *Speech Monographs*, XIII (1946), 19–23.

2) Paul Moore, "Intelligibility related to loudness," *Speech Monographs*, XIII (1946), 13–18.

3) J. W. Turkey, "Comparing individual means in analysis of variance," *Biometrics*, 5 (1949), 99–114.

4) *Final Report in Summary of Work on Voice Communication*, OSRD Report No. 5568, 1945, PB No. 1205; U. S. Dept. Commerce, 1946.

5) *An Experimental Comparison of Five Conditions for Voice Communication Training*, Project 20-K-1, Contract N6ori-104, Technical Report No. SDC 104-2-4, Purdue Voice Communication Laboratory, August, 1947.

6) *Voice Communication: Retention of Improved Intelligibility*, Project 20-K-1, Contract N6ori-104, Technical Report No. SDC 104-2-5, Purdue Voice Communication Laboratory, January 1948.

7) *Voice Communication: Retention of Improved Intelligibility II*, Memorandum Report, Project 20-K-1, Contract N6ori-104, Technical Report No. SDC 104-2-9, Purdue Voice Communication Laboratory, April 1948.

8) *Retention of Voice Communication Training after One Year*, Project 20-K-1, Contract N6ori-104, Technical Report No. SDC 104-2-13, Purdue Voice Communication Laboratory, March 1949.

INTER-GRADE COMPARISONS OF WORD FREQUENCIES IN CHILDREN'S WRITING

GERTRUDE HILDRETH

Brooklyn College

A practical alphabetical list indicating the relative frequency with which words are used by children in writing is a requisite for elementary school spelling instruction. The basic assumption underlying the construction of the list devised by Hildreth and Salisbury[1] and of the revision prepared by Parke[2] is that the words commonly used in children's everyday writing should be taught ahead of those less commonly used. The total frequencies for all grades combined as given in the Rinsland Vocabulary[3] were used in the construction of the Hildreth-Salisbury list as well as the New York City list when preliminary correlation studies showed the high degree of correspondence in frequency rank from grade to grade for the commoner words in the total Rinsland list.

It is important to determine the degree of relationship in frequency rank for words of the separate grade lists of the Rinsland Vocabulary and total frequency for all grades combined, because this information would aid in determining whether the total frequencies for particular words are as valid for word selection and sequential gradation of spelling words as the separate grade lists. If so, the separate sub-lists by grades can be disregarded, and the total frequency column in Rinsland furnishes valid information for preparing word lists arranged in frequency levels; separate grade lists for the intermediate and upper elementary grades then become unnecessary.

Another interesting question is the stability in frequency rank from grade to grade of words more commonly used by children in writing in comparison with those less commonly used. For example, does such a common word as 'chair' have a more consistent rank

---

[1] Gertrude Hildreth. "Spelling as a language tool." *Elementary School Journal*, XLVIII, (September, 1947) 33–40.

[2] Margaret Parke. *A Manual to Guide Experimentation With Spelling Lists A, B, and C*. New York City Board of Education, 1951.

[3] Rinsland, H. D. *A Basic Vocabulary of Elementary School Children*. New York: The Macmillan Co., 1945.

REFERENCES

1) Harry M. Mason, "The effect of very loud speech signals upon intelligibility," *Speech Monographs*, XIII (1946), 19–23.

2) Paul Moore, "Intelligibility related to loudness," *Speech Monographs*, XIII (1946), 13–18.

3) J. W. Turkey, "Comparing individual means in analysis of variance," *Biometrics*, 5 (1949), 99–114.

4) *Final Report in Summary of Work on Voice Communication*, OSRD Report No. 5568, 1945, PB No. 1205; U. S. Dept. Commerce, 1946.

5) *An Experimental Comparison of Five Conditions for Voice Communication Training*, Project 20-K-1, Contract N6ori-104, Technical Report No. SDC 104-2-4, Purdue Voice Communication Laboratory, August, 1947.

6) *Voice Communication: Retention of Improved Intelligibility*, Project 20-K-1, Contract N6ori-104, Technical Report No. SDC 104-2-5, Purdue Voice Communication Laboratory, January 1948.

7) *Voice Communication: Retention of Improved Intelligibility II*, Memorandum Report, Project 20-K-1, Contract N6ori-104, Technical Report No. SDC 104-2-9, Purdue Voice Communication Laboratory, April 1948.

8) *Retention of Voice Communication Training after One Year*, Project 20-K-1, Contract N6ori-104, Technical Report No. SDC 104-2-13, Purdue Voice Communication Laboratory, March 1949.

INTER-GRADE COMPARISONS OF WORD FREQUENCIES IN CHILDREN'S WRITING

GERTRUDE HILDRETH

Brooklyn College

A practical alphabetical list indicating the relative frequency with which words are used by children in writing is a requisite for elementary school spelling instruction. The basic assumption underlying the construction of the list devised by Hildreth and Salisbury[1] and of the revision prepared by Parke[2] is that the words commonly used in children's everyday writing should be taught ahead of those less commonly used. The total frequencies for all grades combined as given in the Rinsland Vocabulary[3] were used in the construction of the Hildreth-Salisbury list as well as the New York City list when preliminary correlation studies showed the high degree of correspondence in frequency rank from grade to grade for the commoner words in the total Rinsland list.

It is important to determine the degree of relationship in frequency rank for words of the separate grade lists of the Rinsland Vocabulary and total frequency for all grades combined, because this information would aid in determining whether the total frequencies for particular words are as valid for word selection and sequential gradation of spelling words as the separate grade lists. If so, the separate sub-lists by grades can be disregarded, and the total frequency column in Rinsland furnishes valid information for preparing word lists arranged in frequency levels; separate grade lists for the intermediate and upper elementary grades then become unnecessary.

Another interesting question is the stability in frequency rank from grade to grade of words more commonly used by children in writing in comparison with those less commonly used. For example, does such a common word as 'chair' have a more consistent rank in each separate grade than a less commonly used word such as 'chapter'. For younger children in Grades III and IV teachers are chiefly concerned about a short list of high frequency words.

---

[1] Gertrude Hildreth. "Spelling as a language tool." *Elementary School Journal*, XLVIII, (September, 1947) 33–40.

[2] Margaret Parke. *A Manual to Guide Experimentation With Spelling Lists A, B, and C.* New York City Board of Education, 1951.

[3] Rinsland, H. D. *A Basic Vocabulary of Elementary School Children.* New York: The Macmillan Co., 1945.

Method.—The problem was attacked by making a random sampling of one hundred words from the Hildreth-Salisbury List, Levels I through VIII, Level I consisting of the most commonly used words according to total frequency such as 'my', 'very', 'our'; Levels VI, VII and VIII consisting of relatively infrequently used words such as 'zone', 'machinery', 'quart', and then comparing the deviations in rank for these one hundred words in the separate Rinsland grade lists for Grades III, IV, VI and VIII. Omitted from the one hundred words were abbreviations, two-word combinations, and proper names.

Words in the Rinsland grade levels I and II were omitted from these comparisons because these grades obviously present a different problem in spelling word selection than the higher years. Grades V and VII were omitted because it seemed unnecessary to do the computations for each one of the separate upper elementary grades.

In order to determine the intergrade relationships for the commoner words the same comparisons in terms of deviation in frequency rank were made for the forty commonest words in the sampling of one hundred according to Rinsland total frequency. These forty words correspond roughly to the commonest 1800–2000 words used by elementary school children. The entire hundred words correspond approximately to the commonest 4500–4800 words used by children in their writing. Intergrade deviations for the forty commonest words were not computed because the trends are evident from the comparisons with the respective grade frequencies and total frequency for the hundred-word list.

Frequency distributions were made for these deviations in rank without respect to sign in step intervals of 1 and the medians for all these distributions were computed. Table I shows the distributions of the deviations, the medians for each set of deviations, and the range in deviation for each set of comparisons. In Table I, T stands for Total Frequency. The grade designations at the head of each column, e.g., 3–4, 3–6, etc. mean Grades III and IV, III and VI, and so on. The medians were computed on ungrouped distributions of deviations.

The following deviations in frequency rank for each of the hundred words were computed: Grades III and IV, III and VI, III and VIII, III and total; IV and VI, IV and VIII, IV and total; VI and VIII, VI and total, VIII and total. For the forty high

TABLE I. DISTRIBUTIONS OF DEVIATIONS IN RANK AT VARIOUS GRADE LEVELS FOR SELECTIONS OF WORDS IN THE HILDRETH-SALISBURY LIST
(medians computed on ungrouped data)

Deviation	Data for 100 words in the total list										Data for 40 words of Highest total frequency			
	Gr. 3–4	Gr. 3–6	Gr. 3–8	Gr. 3-T	Gr. 4–6	Gr. 4–8	Gr. 4-T	Gr. 6–8	Gr. 6-T	Gr. 8-T	Gr. 3-T	Gr. 4-T	Gr. 6-T	Gr. 8-T
50+		1	1											
48														
46														
44		1	1											
42					1	1								
40			2			1								
38			1			1								
36	1	1	1					2						
34			4	1										
32	2	1				2		2		1				
30	1	1	2	1	1	2		1		1				
28		3			2	2		1		1				1
26		3	1			2		2		1				1
24	1	1	2	2	2	1				1				
22		1	6	3	4	4	3	3	2	1			1	1
20	1	3	8	2	3	2		2	2	2				
18	3	2	4	2	3	4		4	2	3				1
16	5	3	1	3	5	1	3	9	2	6	1			1
14	3	8	9	9	7	7	2	4	3	3	1		1	1
12	8	4	7	5	6	6	8	7	3	3		2	1	
10	7	10	6	5	11	7	8	5	7	7	1	1	2	3
8	7	10	13	5	3	6	7	12	11	8	1	2	5	3
6	12	9	9	14	10	10	6	10	11	15	7	1	5	6
4	12	11	4	16	13	11	14	8	18	11	11	8	8	6
2	16	13	6	17	18	16	19	13	19	20	7	10	7	9
0	21	14	12	15	11	14	30	15	20	17	11	16	10	7
N	100	100	100	100	100	100	100	100	100	100	40	40	40	40
Range	0–37	0–52.5	0–50	0–34	0–43	0–42	0–23	0–37	0–22	0–30	0–16	0–13	0–22	0–28
Median	6.17	9.14	12.0	6.2	7.3	7.83	4.12	8.5	5.14	6.3	4.4	3.0	4.6	5.2

frequency words the following deviations in frequency rank were computed: III and total; IV and total, VI and total, VIII and total.

Results.—Table I shows the median deviations in rank for the frequency order of word usage in the comparisons made among the respective grades and total frequency to be as follows:

Grades III and IV, 6.17; III and VI, 9.14; III and VIII, 12.0, III and total, 6.2; Grades IV and VI, 7.3, IV and VIII, 7.8, IV and total, 4.12; Grades VI and VIII, 8.5, VI and total, 5.14, VIII and total, 6.3. These results indicate that the different words are used with about the same relative frequency in the upper elementary grades. There is so little difference in the deviations for word frequency in Grades IV and VI, IV and VIII, and VI and VIII, (7.3, 7.8, and 8.5) and these deviations are so small as to indicate that separate frequency lists for these grades need not be constructed.

The Grade IV frequency ranks are closest to the total, possibly because Grade IV represents a halfway point in the range of grades from I to VIII and would be less influenced by the brevity of the Grade III list or the wide range of the Grade VIII words.

Inspection of the data for Grade III shows that although the deviations in comparison with the separate grades is lower for the adjacent Grade IV than for Grades VI and VIII, nevertheless the comparison between Grade III and total frequency shows virtually the same deviation as that between Grade VIII and total, both of these deviations being slightly higher than for Grades IV and total, and Grade VI and total. These findings suggest that the Grade III list is somewhat more restricted, less representative than the Grade IV and Grade VI lists, as would be expected from a cursory examination of the writings of third-graders. By Grade IV children are 'hitting their stride' in writing, and using almost as wide a range of the most common words with about the same frequency as they will use them later on in the higher grades.

Our interest in Grade III words for purposes of teaching spelling is confined chiefly to words that are commonest among 2000 in English usage, rather than the entire range of which the hundred words are representative.

The comparisons for the Grade VIII frequencies seem a bit out of line as do the comparisons for Grade III, but for a different reason. The Grade VIII list appears to be more influenced than the lists for Grades IV and VI by words used in formal school theme writing and formal correspondence.

The smaller deviations for each of the separate grade comparisons with total frequency are due in part to intercorrelation, since the total frequency contains the frequency tabulations for the particular grade in question as well as for all others; but the lower deviations are also due in part to the greater stability of the total frequency list which combines frequency counts for eight separate grades and hence is many times larger than any of the separate grade frequencies.

Some words show wide deviations in frequency rank from grade to grade, some relatively small; for example, the average deviation in all intergrade comparisons for the word 'got' is 1; for 'dear', 4; for 'handkerchief', 16; for 'April', 29.

For the forty words of highest total frequency the median deviations are as follows: Grade III and total, 4.4; IV and total, 3.0; VI and total, 4.6; VIII and total, 5.2.

These deviations compared with those for the total list of one hundred words prove what one might expect—that the frequency rank order for the commonest words is more consistent grade for grade than the order for less common words. The egocentric 'my' and 'our' come out on top no matter what the grade level of children using the words.

The less frequently a word is used by elementary school children according to total frequency count, the less consistent the rank from grade to grade, in general. This conclusion was to be expected, because uncommon words are in a sense specialized, e.g., 'machinery', 'quart', 'office', 'April'. Who can predict with certainty just when and where these words will be used in writing by anyone?

The wide deviations certain words show ('chance' has a deviation of 50 for Grades III and VIII; 'April', a deviation of 52.5 for Grades III and VI; 'office', a deviation of 44 for Grades III and VI) are not entirely due to chance. These are obviously less frequently used words in general and words of a more highly specialized character. Inspection of the original tally sheets proves that the words of highest frequency among these forty commonest words show more consistency in rank than the total forty commoner words in the selection of one hundred.

For the forty commonest words, Grade III now appears to be more in line with the other grades. (Note the deviations for Grade

III and total for the hundred words and Grade III and total for the forty high frequency words.) For the forty words, the Grade III median deviation compared with the total frequency is slightly under that for Grades VI and VII, but still higher than for Grade IV. This finding is more significant than the results for the hundred word comparisons because these forty words represent the commonest 1800–2000 words in frequency of use. This is the maximum vocabulary with which any spelling work in the primary grades need be concerned.

In evaluating all these comparisons some allowance must be made for statistical unreliability of the differences between medians due to errors of sampling.

These results tend to indicate that the separate grade frequencies as given in Rinsland probably have no more validity for 'grading' spelling words in order of frequency of use in writing than the Rinsland total frequency list for all grades. There can be little question of this conclusion for the 2000 or so commonest words used in the English language. In the long run, the total frequency list representing the largest number of cases is the most stable, valid, and reliable for the sequential gradation of spelling words. The highly specialized nature of word usage above the 2000 or the 2500 word limit suggests the difficulty of determining valid grade placement of these words.

Common sense suggests that word selection for spelling in the primary grades does present a somewhat different problem than word usage in Grades IV and above. No one would advocate giving the younger children words for study arbitrarily selected from a frequency rank list no matter how consistently the commonest words are used at all grade levels.

Here is a promising field for further research. Additional studies should be made to determine the inter-grade relationships in frequency rank for a larger sampling of words in each of the separate grade levels of the Rinsland List and throughout the entire range of the words listed there. More extensive studies should be made of the relative correspondence in frequency rank of common and uncommon words, of special terms, proper names, and so on. It would be interesting also to discover the relationship that exists when the word usage of children in the elementary grades is compared with the word usage of high school students and adults.

COMPARISON OF PSYCHOLOGY INSTRUCTORS AND NATIONAL NORMS ON THE PURDUE RATING SCALE

A. W. BENDIG

University of Pittsburgh

As was pointed out in a previous report (*2*), student ratings of instructors in introductory psychology can serve two needs: (a) as part of a multiple criterion of teaching competence when combined with other measures of the instructors' teaching behavior, and (b) as a source of information for the instructor in diagnosing his teaching strengths and weaknesses as seen by his students. Before student ratings can be used in an evaluative procedure much more needs to be known about the influence of characteristics of the students (sex, academic achievement, interests, etc.) upon their ratings and about the relationships of such ratings to other measures of teaching competence (supervisor and peer evaluations, objective test scores, speech patterns, etc.). However, student scales can more immediately be used to aid the instructor in discovering and modifying his more obvious deficiencies. Resistance to the use of student rating scales in evaluating the competence of college teachers is both widespread and, in view of the obvious inadequacies and unproven validity of most scales, justifiable at this time. Their usefulness to the instructor as a help in self-diagnosis, however, is more accepted and rests upon a sounder basis.

One necessity in using student scales as diagnostic tools is normative data. For one such set of scales, the Purdue Rating Scale for Instruction (PRSI), percentile norms based upon the mean ratings of two hundred and five college instructors in many different subject matter areas at Purdue University have been provided by Remmers and Baker (*9*). In previous semesters these norms have been utilized in reporting student ratings to psychology instructors at the University of Pittsburgh who requested a student evaluation of their teaching. However, it soon became evident that these norms could not be meaningfully applied to the ratings of our instructors. For example, the mean rating of introductory psychology instructors on PRSI scale 3 (Fairness in Grading) fell at the 90th percentile of the Purdue norms and at the 13th percentile on scale 10 (Stimulating Intellectual Curiosity). These results suggested that either our sample of departmentally homogeneous instructors was not comparable to the heterogeneous group of instructors reported upon by Remmers and Baker, or that University of Pittsburgh students evaluate their instructors somewhat differently than do students at Purdue University.

The first of the above suggestions appears to be the more tenable. All of our instructors are employed by the same department and have similar professional interests and training. In contrast, the Purdue norms are based upon ratings of instructors from many different departments and disciplines. In other reports it has been suggested that inter-disciplinary differences in the subject matter taught may differentially affect student ratings of instructors (*5*) and even intra-department course content variations may be important (*4*). In addition to the subject matter homogeneity, most of our instructors tend to be younger than the average age of university instructors and to have had less teaching experience. Data reported by Goodhartz (*7*) suggest that younger instructors tend to be rated more favorably by students than do older staff members. Descriptions of the professional characteristics of our instructors can be found in previous articles (*2*, *4*). As to student differences between Purdue and Pittsburgh, population data on students enrolling in introductory psychology at the University of Pittsburgh led to the conclusions that "the day-time students fit the picture of the average college student at most institutions" (*2*, p. 169).

The obvious solution to our norms problem was to formulate our own norms based upon the data available upon our instructors. However, the non-comparability of our data raised a question as to the reliability of the PRSI scales when used with our sample. Remmers and Baker (*9*, p. 4) report high reliability for the scales based upon one hundred and fourteen instructors each rated by varying numbers of students. Before our proposed norms could be used, the reliability of the mean ratings based upon our sample had to be investigated.

PROCEDURE

During the Spring and Fall semesters of 1951 eleven instructors taught introductory psychology to daytime undergraduate students. Of the eleven, ten were male. They varied in academic rank from the lecturer to associate professor levels. All had had a minimum of one year of college teaching before instructing in this course. The PRSI ratings were collected two or three class periods before the final examination by members of the department other than those teaching in this course. Quantification of the ratings followed the procedure described in the PRSI Manual (*9*, p. 12). Only data from the first ten scales of the PRSI were used since the primary need for normative data was on the scales referring to instructor characteristics. The remaining sixteen scales of the PRSI refer to course characteristics. The number of students rating each of the eleven instructors varied from seventeen to ninety-eight.

The reliability of the average ratings from each scale was computed by two methods: (a) the intraclass reliability estimate described by Ebel (*6*), and (b) the generalized reliability formula developed by Horst (*8*). When unequal numbers of students rate each instructor these two estimates may differ somewhat because of the different weighting applied to the average rating of each instructor. The average rating of a given instructor contributes to the intraclass estimate in proportion to the number of students rating that particular instructor, whereas the generalized formula weights the average rating of each instructor equally in determining scale reliability, regardless of the differing number of raters from instructor to instructor.

The obtained results in terms of the mean, median, standard deviation, and the two reliability estimates of each scale can be found in Table 1. In addition, comparable data from the Purdue norm group are also given.

It is to be noted in Table 1 that PRSI scale 5 (Presentation of Subject Matter) was the most discriminating scale with our homogeneous group of instructors and scale 3 (Fairness in Grading) appears the poorest. This latter result is most probably attributable to the well-structured departmental grading policy which was followed by all instructors and permitted little individual instructor influence over the grade achieved by each student (*1*, *5*). In addition, the mean rating for scale 3 was the highest of the ten scales, indicating the students approved of the objective grading system used and recognized that there was little difference between instructors as to this characteristic. Scales 1, 5, 7, 9, and 10 appear to be quite reliable, while the reliability estimates of the remaining five scales suggest that less reliance should be placed upon their evaluation of our introductory psychology instructors. Previous data have indicated a generally negative skew to the distribution of ratings on these scales (*5*,*9*, p. 10) and a comparison of the means and medians in Table 1 shows this to be true, in varying degrees, for eight of the ten scales. Revising the wording of the scale categories to reduce skewness in the five scale showing unsatisfactory reliabilities would probably improve their discrimination as was found in modifying similar scales (*3*).

TABLE 1.—NORMATIVE DATA ON MEAN RATINGS OF ELEVEN INTRODUCTORY PSYCHOLOGY INSTRUCTORS ON THE PURDUE RATING SCALE FOR INSTRUCTION (AVERAGE NUMER OF RATERS = 42.8) AND COMPARISON WITH PURDUE NORMS

Purdue Scale	Psychology Norms					Purdue Norms	
	Mean of Means	Median of Means	Standard Deviation of Means	Reliability		Median of Means	Reliability: Generalized
				Intra-class	Generalized		
1	83.5	86.5	7.7	0.91	0.91	90	0.92
2	89.0	89.5	3.6	0.82	0.61	87	0.92
3	91.9	92.3	3.9	0.58	0.54	86	0.86
4	85.5	86.4	4.7	0.81	0.64	85	0.91
5	66.6	66.7	13.7	0.96	0.93	79	0.93
6	79.6	81.5	6.6	0.88	0.75	83	0.90
7	77.7	80.0	8.6	0.91	0.90	84	0.92
8	79.2	77.6	5.8	0.76	0.65	83	0.92
9	87.7	89.4	7.3	0.92	0.91	92	0.94
10	67.8	67.3	8.7	0.92	0.84	78	0.91

Comparing the Horst reliability estimates reported by Remmers and Baker (*9*, p. 4) with the present data indicates a generally higher reliability of the scales as used with the Purdue sample. This lessened reliability is attributable to two factors: (a) the previously mentioned departmental control over the testing and evaluative procedures used by the instructors, thus reducing inter-instructor variability in the eyes of the students, and (b) the greater homogeneity of both course and instructional content between instructors, since the data were obtained from a limited sample of instructors teaching a specific course in a single university department. In view of the tremendous influence of these factors in reducing instructor heterogeneity in our sample, the fact that PRSI scales

1, 5, 7, and 9 yielded reliabilities quite comparable to those found with the heterogeneous Purdue sample is a tribute to the discriminatory ability of both students and scales.

Two suggestions are offered based upon the above results. First, that serious consideration be given to revising the wording of PRSI scale categories to achieve less skew in the distribution of ratings. Secondly, ratings from a fewer number of independent 'scores' for each instructor and evaluations that would be more reliable. A suggested line of attack here might be the factor analytic approach reported by Smalzreid and Remmers (*10*).

In conclusion, these data suggest that the Purdue norms be used cautiously in evaluating the mean PRSI ratings for any group of instructors. Variations in subject matter area and in characteristics of the course taught may seriously invalidate both the percentile norms and scale reliabilities reported in the PRSI Manual. These norms should be used only when an obtained sample of ratings can be shown to be comparable to the Purdue data.

BIBLIOGRAPHY

1) A. W. Bendig. "The reliability of letter grades." *Education & Psychological Measurement* 13, 1953, 311–321.

2) ——— "The use of student rating scales in the evaluation of instructors in introductory psychology." *Journal of Educational Psychology*, 43, 1952, 167–175.

3) ——— "A statistical report on a revision of the Miami Instructor Rating Sheet." *Journal of Educational Psychology*, 43, 1952, 423–429.

4) ——— "A preliminary study of the effect of academic level, sex, and course variables on student rating of psychology instructors." *Journal of Psychology*, 43, 1952, 21–26.

5) ——— "The effect of level of achievement upon students' instructor and course ratings in introductory psychology." *Educational & Psychological Measurement* 13, 1953, 437–448.

6) R. L. Ebel. "Estimation of the reliability of ratings." *Psychometrika* 16, 1951, 407–424.

7) A. S. Goodhartz. "Student attitudes and opinions relating to teaching at Brooklyn College." *School and Society*, 68, 1948, 345–349.

8) Paul Horst. "A generalized expression for the reliability of measures." *Psychometrika*, 14, 1949, 21–31.

9) H. H. Remmers, and P. C. Baker. *Manual of Instructions for the Purdue Rating Scale for Instruction.* Lafayette, Indiana: Division of Educational Reference, Purdue University, 1952.

10) N. T. Smalzreid and H. H. Remmers. "A factor analysis of the Purdue Rating Scale for Instruction." *Journal of Educational Psychology*, 34, 1943, 363–367.

BOOK REVIEWS

Higher Education in the Forty-eight States: A Report to the Governors' Conference. Chicago: The Council of State Governments, 1952, pp. 317. $5.00.

At the 1952 meeting of the Governors' Conference the Council of State Governments reported on an extensive study of higher education in the forty-eight states. The present volume presents this report with five chapters of interpretive text and over one hundred pages of detailed tables. The text is written in a style suited to describe higher education in this country to governmental officials who are not educational experts. At the same time, it contains a wealth of interpretation and extensive statistics of very great value to the specialist in educational administration.

The five chapters deal successively with the history of American higher education, programs in a broad sense but not including internal curricular or administrative problems, finances from the point of view of expenditures and income, and, finally, organization in the sense of the governmental controls of public institutions and the place of such institutions in the state governmental structure.

Comparisons among states in relation to the support of higher education are facilitated by a technique of determining for the states a per mill rate that each state has in relation to the total for the continental United States. The states vary widely in absolute and ratio figures on all variables considered. The proportion of state support for higher education has varied over the years. Although absolute amounts have increased, the percentage of total income from state sources has decreased from about thirty-five per cent in 1918 to twenty-seven per cent in 1950.

In the last chapter, on organization, there is an enlightening analysis of the part played by legislative and administrative officials in the control of state educational institutions, and of the legal and practical authority of institutional boards of trustees. The variation here follows a pattern of differences among the states which is evident in almost all aspects of state government.

At no place do the authors of this report attempt evaluative comparisons. Their function is to describe the status quo and to indicate something of the history which has preceded it. This task has been done extremely well. The basic data are here available for educational or governmental evaluation. C. M. Louttit

University of Illinois

Miriam Forster Fiedler. *Deaf Children in a Hearing World: Their Education and Adjustment.* New York: The Ronald Press Co., 1952, pp. 320. $5.00.

For two hundred years the education of deaf children has been one of the most highly specialized and most controversial areas in education. Since the time of Rudolph Pintner many psychologists have been interested in the intelligence, personality, and adjustment of deaf persons who are so obviously in an environment which they often share in a very limited fashion because of barriers of communication. If we add to the number of educators and psychologists the parents of the more than twenty thousand hearing-handicapped children now in special schools and classes, as well as the puzzled, often frustrated or panic stricken parents of preschool age deaf children, it will total a goodly number of potential readers for any publication with the promising title listed above.

While all of these readers could profit from a perusal of the book, the generalizations implied in the title are somewhat misleading. The book is more precisely the report of a four-week session at the Vassar Summer (1949) Institute for Family and Community Living, attended by eleven young (two and one-half to nine years) hearing handicapped (thirty-seven to eighty-five db loss) children and their parents. The children spent the greater part of the time with hearing children of similar age groups, receiving special help in speech, speech reading, and auditory training. Most of these children could be reached with amplified sound in the speech range and will probably be educated as hard of hearing rather than as deaf. Audiograms are presented for all, but as the author states (p. 19) they are only roughly comparable. A section on the validity and reliability of hearing tests, the EEG and GSR techniques would have added to the scientific aspects of the book.

The discussion on 'sense training' fails to bring out the real purpose of these activities in the first weeks of a small deaf child's schooling. Sense training is not an end in itself, but something which the deaf child can do successfully in coöperation with an adult. This coöperation is a necessary step if speech and language training are to begin, for even though it may be an "antiquated theory" (p. 9) there is historically no record of a deaf child's developing conversational speech and language without formal training. The teaching profession would welcome evidence that it could happen.

These children were young, and the young deaf child will never again be so like his hearing counterpart. The plea is for 'adjustment', rather than for emphasis on communication skills. As the deaf child grows into adolescence and adulthood, what is going to be the measure of his adjustment? Surely his adequacy in speech and language will play an important part. This is the old dilemma of the deaf child's education, and this book does not give a complete solution. The things that the child needs for good adjustment at sixteen or sixty are the things which the hearing child learns without effort and involuntarily before he is six, but which the truly deaf child must learn through voluntary attention to visual cues at any age.

This book is not a report of a controlled experiment, but a sincere attempt to see what would happen under the circumstances. We may assume that staff members not previously acquainted with children with severe hearing losses learned a great deal about such children—a highly desirable outcome in public relations and better understanding of deaf children.

ELOISE KENNEDY
University of Illinois

LLOYD A. JEFFRESS. *Cerebral Mechanisms in Behavior*. The Hixon Symposium. New York: John Wiley and Sons, Inc., pp. 311.

Cerebral Mechanisms in Behavior is a volume containing the content of a symposium on a topic held at California Institute of Technology during the week of September 20 to September 25, 1948. The symposium was sponsored by the Hixon Fund. Lloyd A. Jeffress, who writes the preface in the volume, was selected by the institute to spend the year 1947 in research and audition and help organize the 1948 symposium. The selection of contributions as well as contributors will interest people in the field. They include contributions by John von Neumann on the general and logical theory of automata; an interesting and well-considered chapter by Warren S. McCulloch called "Why The Mind Is In The Head;" a chapter on the problem of serial order in behavior by K. S.

Lashley; a chapter by Heinrich Klüver on functional differences between the occipital and temporal lobes; a chapter by Wolfgang Köhler on relational determination in perception, and a chapter by Halstead on brain and intelligence. Like the atomic scientists, some of the contributors in this volume too have a good sense of perspective, a realization of the significance of what they do, a sense of limitation of it, and even at times a sense of humor about it.

There are some very rich give-and-takes in the volume which is exceedingly well edited and in good taste. The chapter that will particularly interest psychologists is the contribution by Henry W. Brosin on the symposium from the viewpoint of a clinician. Emphasized by Brosin in this chapter is the need for learning a common vocabulary by living and working together with the assumption that reading and experimentation in isolation are inadequate. Brosin's impression is "that the logic of determinism and the pragmatic study of relations and operations are adequate scientific working methods for the skeptical clinician who deals necessarily with poorly defined complex conditions, even though he himself has not been able to create a useful vocabulary for his needs." Brosin ends his contribution with a reference to Boring's often quoted comment that psychology lacks a great man. As far as he is concerned, Freud is that man, and the exploitation of the free-association techniques and the inquiry into the meaning of psychodynamic social patterns can keep us occupied productively for many years. He suggests in his last sentence that psychology may find its great man in the person who can utilize the Freudian concepts and bring them into closer approximation with the ideals of Wundt. The impression of the reviewer is that this conclusion represents a misinterpretation of present observable trends and that the promise of growth appears to be in the application of experimental methods in the exploration and validation of materials and concepts from psychoanalyses, cultural anthropology, and sociology.

The volume contains the well-considered deliberations of men who know what they are talking about and how to write about it, and should prove profitable reading to psychologists and people from the biological sciences interested in the topics of this symposium.

H. Meltzer

Psychological Service Center
St. Louis, Missouri

PUBLICATIONS RECEIVED

James Deese. *The Psychology of Learning*. New York: McGraw-Hill Book Co., 1952, pp. 398. $5.50.

Leonard W. Ferguson. *Personality Measurement*. New York: McGraw-Hill Book Co., 1952, pp. 457. $6.00.

Miriam Forster Fiedler. *Deaf Children in a Hearing World*. New York: The Ronald Co., 1952, pp. 320. $5.00.

Robert Herndon Fife and Herschel T. Manuel. *The Teaching of English in Puerto Rico*. San Juan, P. R.: Department of Education Press, 1951, pp. 410.

Lawrence K. Frank, Ross Harrison, Elisebeth Hellersberg, Karen Machover, and Meta Steiner. *Personality Development in Adolescent Girls*. New Orleans: Society for Research in Child Development, Inc., Louisiana State University, 1953, pp. 316. (paper)

Ernest Gardner. *Fundamentals of Neurology*. Philadelphia: W. B. Saunders Co., 1952, pp. 359.

Frank A. Geldard. *The Human Senses*. New York: John Wiley & Sons, Inc., 1953, pp. 365. $5.00.

Edward B. Greene. *Measurements of Human Behavior*. Revised Edition. New York: The Odyssey Press, 1952, pp. 790. $4.75.

William Griffiths. *Behavior Difficulties of Children as Perceived and Judged by Parents, Teachers, and Children Themselves*. Minneapolis: The University of Minnesota Press, 1952. $3.00.

Bernard H. Hall and Others. *Psychiatric Aide Education*. New York: Grune & Stratton, 1952, pp. 168. $5.75.

Ira J. Hirsh. *The Measurement of Hearing*. New York: McGraw-Hill Book Co., 1952, pp. 364. $6.00.

Sylvan Hoffman, Editor. *News of the World: A History of the World in Newspaper Style*. New York: Prentice-Hall Inc., 1953, 52 four-page units. $3.72.

Hubert S. Howe, Chairman. *Conferences on Drug Addiction among Adolescents*. New York: The Blakiston Co., 1953, pp. 320. $4.00.

William H. Ittelson. *The Ames Demonstrations in Perception: A Guide to Their Construction and Use*. Princeton, N. J.: Princeton University Press, 1952, pp. 88. (paper) $4.00.

Arthur T. Jersild. *In Search of Self*. New York: Bureau of Publications, Teachers College, Columbia University, 1952, pp. 145. $2.75.

A. M. Jordan. *Measurement in Education*. New York: McGraw-Hill Book Co., 1953, pp. 533. $5.25.

Charles Koch. *The Tree Test: The Tree-drawing Test as an Aid in Psychodiagnosis*. New York: Grune and Stratton, 1952, pp. 87.

Raymond G. Kuhlen and George G. Thompson, Editors. *Psychological Studies of Human Development*. New York: Appleton-Century-Crofts, Inc., 1952, pp. 533. (paper) $3.50.

Oliver L. Lacey. *Statistical Methods in Experimentation: An Introduction*. New York: The Macmillan Co., 1953, pp. 249. $4.50.

Paul H. Landis. *Adolescence and Youth: The Process of Maturing*. Second Edition. New York: McGraw-Hill Book Co., 1952, pp. 461. $5.00.

Robert Lindner. *Prescription for Rebellion*. New York: Rinehart & Co., Inc., 1952, pp. 305. $3.50.

E. F. Lindquist. *Design and Analysis of Experiments in Psychology and Education*. Boston: Houghton Mifflin Co., 1953, pp. 393. $6.50.

Olof Magne. *Perception and Learning: Experimental Studies in the Similarities and Dissimilarities between Perception and Learning Tests*. Uppsala, Sweden: Appelbergs Boktryckeri AB, 1952, pp. 228. (paper) 25 kronor.

Norman R. F. Maier. *Principles of Human Relations: Applications to Management*. New York: John Wiley & Sons, Inc., 1952, pp. 474. $6.00.

William E. Martin and Celia Burns Stendler. *Child Development: The Process of Growing Up in Society*. New York: Harcourt, Brace & Co., 1953, pp. 519.

Paul Maslow. *Powers of the Mind*. Volume II. of The Life Science. Brooklyn, N. Y. (16 Court St.): The Author, 1952, pp. 288. (paper) $3.50.

Florence McGehee. *Please Excuse Johnny*. New York: The Macmillan Co., 1952, pp. 242. $3.50.

ANNE SELLEY MCKILLOP. *The Relationship between the Reader's Attitude and Certain Types of Reading Response.* New York: Bureau of Publications, Teachers College, Columbia University, 1952. $2.75.

ASHLEY MONTAGU. *The Natural Superiority of Women.* New York: The Macmillan Co., 1953. $3.50.

CLARK E. MOUSTAKAS. *Children in Play Therapy: A Key to Understanding Normal and Disturbed Emotions.* New York: McGraw-Hill Book Co., 1953, pp. 218. $3.50.

JAMES L. MURSELL. *Psychology for Modern Education.* New York: W. W. Norton & Co., 1952, pp. 610. $3.90.

C. W. ODELL. *How to Improve Classroom Testing.* Dubuque, Iowa: Wm. C. Brown Co., 1953, pp. 156. (paper) $3.00.

LESLIE PHILLIPS AND JOSEPH G. SMITH. *Rorschach Interpretation: Advanced Technique.* New York: Grune & Stratton, 1953, pp. 385.

FLORENCE B. POWDERMAKER AND JEROME D. FRANK. *Group Psychotherapy: Studies in Methodology of Research and Therapy.* Cambridge: Harvard University Press, 1953, pp. 615. $6.50.

EARLE L. REYNOLDS. *The Distribution of Subcutaneous Fat in Childhood and Adolescence.* Evanston, Ill.: Society for Research in Child Development, Inc., 1952, pp. 189. (paper)

FLORENCE GREENHOE ROBBINS. *Educational Sociology: A Study in Child, Youth, School, and Community.* New York: Henry Holt & Co., 1953, pp. 529. $4.75.

HELEN M. ROBINSON. *Clinical Studies in Reading: II.* Chicago: The University of Chicago Press, 1953, pp. 189. (paper) $3.75.

G. WILSON SHAFFER AND RICHARD S. LAZARUS. *Fundamental Concepts in Clinical Psychology.* New York: McGraw-Hill Book Co., 1952, pp. 540. $6.00.

JOSEPH FRANCIS SHARPE. *The Retention of Meaningful Material.* Washington, D. C.: The Catholic University of America Press, 1952, pp. 66. (paper)

EVERETT L. SHOSTROM AND LAWRENCE M. BRAMMER. *The Dynamics of the Counseling Process.* New York: McGraw-Hill Book Co., 1052, pp. 213. $3.50.

B. F. SKINNER. *Science and Human Behavior*. New York: The Macmillan Co., 1953, pp. 461. $4.00.

MARJORIE B. SMILEY. *Intergroup Education and the American College*. New York: Bureau of Publications, Teachers College, Columbia University, 1952, pp. 212. $4.25.

L. SZONDI. *Experimental Diagnostics of Drives*. Translated by Gertrude Aull. New York: Grune & Stratton, 1952, pp. 254. $13.50.

ARTHUR G. TANSLEY. *Mind and Life. An Essay in Simplification*. London: George Allen & Unwin Ltd. (New York: John DeGraff, Inc.), 1952, pp. 171. $3.50.

CHARLES B. THOMPSON AND ALFREDA P. SILL. *Our Common Neurosis: Notes on a Group Experiment*. New York: Exposition Press, 1953, pp. 210. $3.50.

ARTHUR E. TRAXLER, ROBERT JACOBS, MARGARET SELOVER, AND AGATHA TOWNSEND. *Introduction to Testing and the Use of Test Results in Public Schools*. New York: Harper & Bros., 1953, pp. 113. $2.50.

PHILIP E. VERNON. *Personality Tests and Assessments*. London: Methuen & Co., Ltd., 1953, pp. 220. 18s.

W. EDGAR VINACKE. *The Psychology of Thinking*. New York: McGraw-Hill Book Co., 1952, pp. 392. $5.50.

LOUIS B. WEHLE. *Hidden Threads of History: Wilson through Roosevelt*. New York: The Macmillan Co., 1953, pp. 300. $4.00.

HEINZ WERNER AND EDITH KAPLAN. *The Acquisition of Word Meanings: A Developmental Study*. Evanston, Ill.: Society for Research in Child Development, Inc., 1952, pp. 120. (paper)

CARL A. WHITAKER AND THOMAS P. MALONE. *The Roots of Psychotherapy*. New York: The Blakiston Co., 1953, pp. 236.

JOHN W. M. WHITING AND IRVIN L. CHILD. *Child Training and Personality: A Cross-cultural Study*. New Haven, Conn.: Yale University Press, 1953, pp. 353. $5.00.

FRANCES M. WILSON AND MORRIS KRUGMAN. *Studies of Student Personnel*. Albany, N. Y.: The University of the State of New York, 1951, pp. 197. (paper).

DALE WOLFLE, Chairman. *Improving Undergraduate Instruction in Psychology*. New York: The Macmillan Co., 1952, pp. 60. $1.25.

MAURICE D. WOOLF AND JEANNE A. WOOLF. *The Student Personnel Program: Its Development and Integration in the High School and College*. New York: McGraw-Hill Book Co., 1953, pp. 416. $5.00.

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 44 December, 1953 Number 8

THE VALUE CONCEPT IN EDUCATIONAL PSYCHOLOGY[1]

WM. CLARK TROW

School of Education

University of Michigan

Not long ago I asked a graduate student in social psychology what his subject teaches about value judgments. Without hesitation he smilingly replied with the genotype of all value judgments: "They're bad!" Many social scientists have tried to follow the lead of the natural scientists in consigning value to the limbo of lost souls that never will be missed. And they have been aided and abetted by non-scientific scholars, both religious and secular, who have made *ex cathedra* pronouncements to the effect that science can tell 'how' to do things, but not 'what' to do. Even some who are willing to use science to discover the most efficient means to attain the ends about which there is general agreement will not accept its services to determine what ends should be sought. This position is taken, for example, by Fiegel (7) who states that: "It is one thing to *de*scribe by means of declarative statements what *is* the case, or to predict what will (probably) be the case if certain conditions are fulfilled; it is another thing to *pre*scribe by means of overtly or covertly imperative sentences . . . what *ought* to be done."

TRADITIONAL DUALISM BETWEEN FACT AND VALUE

It does not lie within the scope of this paper to formulate a reply to this point. Suffice it to say that the traditional dualism on the basis of which facts have been accepted as the proper realm of science, and values excluded, has in recent years been vigorously assailed by philosophers and social scientists alike (*13*, *15*, *16*, *25*). The former, influenced to no small degree by Dewey (*5*) are represented among others by Geiger (*8*) who has stated his views as follows: "It is a little surprising then, that one of the most conspicuous (and mischievous) cultural hang-overs still plagues social science, just as it haunts natural science and philosophy. I refer to the antique dualism between fact and value . . . What are the alternatives? What are the substitutes for scientific inquiry in the handling of human values? . . . The answers can be found in brute power or in mystical illumination, in retreat to The Church (whichever one) or in esoteric obscurantism."

---

[1] Presidential address, Division of Educational Psychology of the American Psychological Association, September, 1953.

The psychiatrist, Franz Alexander (*1*) has written: "The discussion of this topic appears to me as much outmoded as a controversy over whether machines heavier than air can rise up against the force of gravity and fly . . . The dichotomy between 'facts and values' is a pseudo-distinction, and the problem of whether values belong to a realm which is beyond the reach of scientific methods is a pseudo-problem."

The anthropologist, Clyde Kluckhohn (*11*) believes that, "No tenet of intellectual folklore has been so damaging to our life and times as the cliché that 'science has nothing to do with values.' If the consideration of values is to be the exclusive property of religion and the humanities, a scientific understanding of human experience is impossible." And he quotes F. S. C. Northrup: "The norms for ethical conduct are to be discovered from the ascertainable knowledge of man's nature, just as the norms for building a bridge are to be derived from physics."

E. L. Thorndike (*26*), in his presidential address before the American Association for the Advancement of Science in 1936, stated that, "Judgments of value are simply one sort of judgments of fact, distinguished from the rest by two characteristics: They concern consequences. These are consequences to the wants of sentient beings. Values, positive or negative, reside in the satisfaction or annoyance felt by animals, persons, or deities."

VALUE PRESSURES ON THE SCHOOLS

In recent years philosophers have given increasing attention to axiology, the study (or science) of values. A few psychologists have addressed themselves to problems of choice or preferential behavior in organisms of higher complexity than the white rat, and some have pondered the nature of the social values of psychologists. Anthropologists have observed and reported culturally-defined and strongly-held beliefs, codes and sanctions, and sociologists and economists have theorized about social and economic values, respectively. But educational psychologists have given scant attention directly to the problem of values and value judgments or seen that value theory, unformulated though it may be, permeates all educational interrelationships. It is my contention that this problem of values extends into the field of educational psychology, since it involves the study of the process by which individuals learn to make choices of means and ends, and of the interrelationships of these choices to each other in the life orientation of the individual.

Although educational psychologists and others directly concerned with educational problems for the most part have been silent on the subject of values, such taciturnity has not characterized the educational views expressed by certain other groups. Churchmen, for example, have of late begun to hurl the charge at the schools that they do not teach 'moral and spiritual values,' and about all the schoolmen could do was to shout back, 'We do so!' and then turn around and ask, 'What are values anyway?' The question was passed on up to the highest educational echelons, and at a conference of the Educational Policies Commission, certain members were unwilling to admit even that good health is a value, so great was the confusion on the subject. And the Commission as a whole could not agree on a statement regarding values freed from the sanctions of theocratic ideology. As a consquence it was decided to issue a monograph making a few suggestions (*6*) and encourage workshops in which teachers would develop their own concepts of moral and spiritual values, though it was pointed out that when spiritual values are defined in terms of theological doctrine, they are properly not a part of the school program.[2]

The ultra-intellectuals of the humanistic persuasion have likewise added their voices to those of the ecclesiastical critics, contending that the function of the schools is to transmit the cultural heritage and give training in the 'intellectual disciplines,' usually with particular emphasis on their own academic specialization. Replies to these gentlemen (*28*) have had to point out, among other things, that there are likewise other values to be considered, inasmuch as their predilection for arbitrarily determined academic standards for all pupils revealed an extremely narrow view even of what knowledge is of most worth. Much more damaging potentially at least have been the attacks of the really hard-boiled ruffians who have employed the most questionable social-action tactics in several communities to win over supposedly democratically-reared citizens to their anachronistic obscurantism.

---

[2] Since the term 'spiritual' has so many different meanings (*27*), it might be better if its use were forbidden in educational writings and speeches, or if authors were required to furnish a definition of their meaning. (It might be going too far to demand an operational definition!)

VALUE CONCEPTS IN EDUCATION

However, the axiological naiveté of many educators and of citizens generally when subjected to such pressures as these does not imply that problems of value are new to education. Scattered and uncoördinated efforts have here and there been successful in building value concepts into the structure of educational theory and practice. These efforts may be referred to briefly as intuitional, empirical, rational, and experimental.

The intuitional efforts are illustrated by the work of the great educational reformers, Rousseau, Pestalozzi, and Froebel, who did much to shift the value orientation of the schools from an authoritarian, society-centered pattern to one emphasizing individual needs and development. The importance of many values that have been traditionally neglected has been emphasized by the Progressive Education group of our own time, but still largely on an intuitional basis.

Values have of necessity been handled empirically by school boards and superintendents, since they are subject to the pressures of public opinion. As a consequence, health values (and some others as well) were forced in, but they had to enter through the back door of after-school, extra-curricular activities. Intellectual values have, of course, always been in good repute, but they have suffered from the enthusiasm both of the instrumental-value, formal disciplinarians, and of the latter-day saints of the aristocratic tradition, the absolute-value devotees of the humanities who believe, properly enough, that the schools should help to transmit the culture of the past, but who have different ideas as to what aspects of the culture should be transmitted. Political or power values have been encountered on the basis of which, at one extreme autocratic methods to develop subservience are advocated, while at the other democratic leadership is encouraged. Economic values have long been stressed from the early New England demand for the three

R's, trade training, and theological preparation, and later through the development of a wide range of trade and professional schools and the present-day emphasis on vocational education and guidance. Esthetic values have been generally neglected, while social values have been governed largely by the punitive theories derived from the tradition of Puritanism and of British legal positivism. The important task of helping young people to find any harmony of values in their life orientation has been left to fall between the mores-bound influence of their homes, the conflicting theologies of the churches, and the not very effective system of required and elective courses.

In contrast with this unsatisfactory empiricism, certain rationalistic hypotheses have been worked out and reflected in educational theory, even if they are not often clearly observable in practice. Evolutionary theory reinforced the intuitional view that nature is right and, therefore, should be followed, and the instinct hypothesis provided a temporary framework for understanding motivation, coupled with earlier motivational theory based hedonistically on pleasantness and unpleasantness which had introduced 'sugar-coating' methods into the schools. A behavioral theory of choice was embodied in Thorndike's satisfiers and annoyers, which seem still to be acceptable if translated to read, "objects and conditions having positive and negative valence." Psychoanalysis, though slow to catch on, is now deluging the journals with a motivation theory based on unconscious conflict (*19*, *10*).

Although Thorndike's reward-and-punishment experiments in the preferential behavior of chicks and kittens, which sired a library full of maze and puzzle-box experiments, are now more than fifty years old, and his 'Right'-'Wrong' experiments on young adults have attained their majority, they are rarely interpreted as having axiological significance. And the same is true of the Gestalt elaboration of perception and the Lewinian emphasis on a power field and psychological barriers. Paper-and-pencil tests of interests and attitudes come still closer to the heart of the axiological problem, while personality inventories, ratings, social-class investigations, opinion surveys, and case studies are right in the middle of it, though the researchers may not realize their location, and the results are largely unsystematic and unrelated.

As a consquence of the welter of intuition, empiricism, theory, and experimentation, educators have actually become well aware of one aspect or another of the problem of values in education, but under another name. They have dealt with it as the problem of educational aims or objectives. They were forced by practical necessity to harmonize in some way the conflicting value claims of various pressure groups, theories, and research studies and the way chosen was to name committees, from time to time, to formulate a comprehensive list of educational objectives to serve as a kind of creed. The 1918 report of the Commission on the Reorganization of Secondary Education (*22*) listed seven 'cardinal' objectives: health, command of fundamental processes, worthy home membership, vocation, civic education, worthy use of leisure time, and ethical character. This value-packed list is still a good one, though lacking in anything that could be called a basic theory, as the term is technically used. That the list was not entirely satisfactory is evident from the fact that many other lists have appeared since, and various techniques have been tried out for formulating new objectives.

A direct attack on the problem of values has been made by relatively few investigators. Semantic difficulties have bedeviled the philosophers in their efforts to agree on the meaning of the traditional axiological concepts (*15*), and as a consequence a spate of new terms has gushed from the pens of the social scientists, particularly, Tolman, Parsons, Murray, and Kluckhohn (*23*). Although Woodruff (*32*) and others have used other lists of values, most of the research studies have employed the Allport-Vernon *Study of Values* (*2*) with understandably meager results since relatively little work has gone into this test since it first appeared in 1931 by way of item analysis, alternate forms, and adaptation to different age and culture groups.

VALUES IN EDUCATIONAL WORK AREAS

If educational psychologists are really to come to terms with the problem of values, they must do so in the three areas, of 'theory', 'research', and 'practice'. The question of values is a significant one in learning theory and in personality theory, as psychologists, social psychologists, sociologists, and cultural anthropologists are coming to realize, but scientists in these fields have not yet really begun to exploit the possibilities of the school as a social institution, or of the child in school as a learning, adjusting organism. If educational psychologists are to assume the responsibility which is theirs, they will include findings from these fields within their province and also the cross-disciplinary concepts that are coming into favor, and no longer be satisfied with a pet list of wants or needs and vague intuitions as to the proper structure of a democratic school system. And if a satisfactory value theory for education is to be built, they will apply what is developing in semiotics, starting perhaps with Morris' (*20*) designative, appraisive, and prescriptive signs. They will distinguish more carefully between the desired and the desirable, and learn to identify and validate their criteria for the latter. More specifically, they will learn to distinguish between the 'I like' of personal taste or interest, the 'I want' of desire, the 'I need' of organic or social demand, the 'I ought' of interiorized sanctions, and the compulsive 'I must'. And they will likewise be able to put these different preferential verbalisms into the second and third person, and use them in the plural as well as the singular, on a rational, objective basis, as a result of impartial inquiry into facts and consequences, instead of deciding what others need and what ought to be done entirely from the narrow frame of reference of personal and group prejudice. They will of necessity analyze the dimensions of value suggested by Kluckhohn (*23*, p. 412 ff.), and translate them into the forms of behavior found in educational situations.

Well-formed scientific theory leads to its embodiment in research. As Hull (*9*) has said, if the actual, dynamic conditions unfold as a theory implies, the theory acquires an increment of verification, and the hypothesis so tested may tentatively be regarded as true; otherwise as false. So far, very little has been done to determine the effectiveness of the means for attaining the commonly held values, save those of intellect, and less still to discover whether there is a gain or even perhaps a harmful loss to other values resulting from the attainment of a valued goal or from the means employed to attain it. A growing research program will presumably result in the development of new and improved research instruments and techniques. Authors of interest inventories will be far more value-conscious, as in the case of the Maller and Glasser *Interest Values Inventory* (*18*) which contains an adaptation of the Allport-Vernon study. Interview techniques will be used more often to supplement the questionnaire, and other techniques will be developed and improved, such as the Bavelas (*3*) method of discovering a child's surrogate or 'sanction figure' for matters about which he is praised or blamed, and White's (*31*) value analysis method for describing qualitative data. When descriptive research has begun to provide information about the value systems of individuals of differing ages in various economic, social, and cultural groups, we can look forward to studies of the effectiveness of specific educational techniques in developing behavior in the direction of objectives agreed upon as desirable, which will lead to sounder knowledge concerning the desirability of these objectives.

In educational practice instructional materials will no doubt be radically modified in order better to adapt to and promote individual and group values. Already the value shift from a curriculum-centered to an individual ability-centered ideal has resulted in studies of readability and the development of reading materials more accommodating to the imperfectly correlated verbal-ability and interest-maturity variables. The schools of the future will have the task of beginning instruction not only where the child is with respect to intellectual values, which has become an educational axiom, but also of beginning where he is with respect to other values as well. Perhaps most important is the necessary change in operational techniques, as illustrated by the emphasis on democratic in contrast with autocratic and *laissez-faire* group atmospheres. Madden (*17*) calls upon us to exorcise the authoritarian elements in both naturalism and supernaturalism and work for the integration and the differentiation of values through creative social action. The work of the teacher and of the educational psychologist in the school, whatever his title, will necessarily be harmonized, for teaching shares with individual and group guidance the task of helping young people to find a desirable value orientation, and to practice behavior in harmony with it.

There is as yet no generally accepted system of value categories owing largely to the fact that values differ in different groups and cultures and hence tend to be 'culture-bound'. However, there are advantages, for some purposes at least, in using a modification of the Allport-Vernon (*2*) list derived from the ideal types of Eduard Spranger (*24*). These categories are more or less familiar and are fairly easily understood, and their universality and high level of generality permit the values of an individual, an institution, and even perhaps a society to be ranked with respect to them, qualitatively (according to subsumed values) or quantitatively ranging from deficiency or deprivation through a middle optimum range, to satiety (at any one time), or to excess. It would be expected that persons and groups will differ within the optimum range, that institutions, like the school, will provide varied opportunities for their members within this range, each person emphasizing in his life pattern the activities that are of value to him, and that the value orientation which he is helped to find will consist of a (to him) satisfactory balance or harmony among these values.[3] Judgments as to value emphasis will be made with as much foresight of consequences as possible. Such a framework used as a kind of check list calls attention to areas of overemphasis and of neglect, where corrective measures can be initiated.

VALUES AND THE SCHOOL PROGRAM

If such a valuistic approach to educational problems as has been suggested is to be built on modifications in theory, research, and practice, it will have to make direct contact with various aspects of the school program. For convenience these aspects may be regarded as 'curriculum', 'method', 'rating', 'guidance', and 'evaluation'.

The curriculum offers the opportunity for the widest use of value criteria (*4*, *30*), as may be illustrated with reference to the Spranger list. This list omits health values which it has been suggested (*29*, *12*, *21*) might properly be added. For the school these would include instruction in health and sanitation and an athletic and physical education program, one that does not conflict with other value activities. The cognitive or intellectual values have been and continue to be emphasized, but attention to other values also is not an indication of anti-intellectualism, the opinions of some of our intellectuals to the contrary notwithstanding. However, the aristocratic, ultra-intellectualistic tradition has had to give ground in order that other life values could be given positions of priority in the curriculum. Attention to esthetic values would call for the development of the now largely neglected art program, both appreciative and creative, with its satisfactions viewed not only as an end in themselves but also as a means to the maximization of other values in the curriculum (*14*). Attention to the political or power values would include appropriate emphasis on mental hygiene, leadership, and pupil self-government, and to economic values on vocational education and training. The social or altruistic values, unlike the others might have no special period in the school day devoted to them, but like them would permeate and enrich the whole curricular program.

---

[3] It would seem that this is at least one of the ends that religion more or less successfully aims to promote, and therefore that Spranger's religious value with its immanent and transcendental mysticism, aside from its esthetic elements, is somewhat parochial and might advisedly be redefined in terms of desirable value orientation.

Good methods can glorify a meager curriculum and poor methods ruin a good one. Following a valuistic approach, curricular activities would be handled in such a way as to take advantage of all the values contained in them without, of course, needlessly distorting them for this purpose. Research studies in learning, achievement, and growth have thrown a great deal of light on the problems of the slow learner in the area of the intellectual values, but we are still rather primitive in our ways of dealing with learning problems in other areas—power, esthetic, and social, for example.

Ratings are made of teachers and of pupils on trait words which do not derive from any systematic personality theory, and which are seldom defined in behavioral terms. They are made with little or no recognition of the standards of conduct being employed, their source, their validity, or even their desirability, and with no knowledge of norms for different age groups, social class groups or any others. It is here, perhaps, that the evils of the implicit and unrecognized value judgment are to be most easily observed. The words 'bad' and 'good' are commonly used by parents, particularly in the area of the social values, with complete self-assurance. And teachers and administrators employ intervening scale points when required to do so. No need to ask, "Good for whom?" or "Good for what?" It is assumed that the school is responsible for the 'good' behavior, not the 'bad', for the 'bad' is evidence of the child's evil nature. Fortunately, many teachers are becoming a little less sure of the ratings they may still be required to make though they may not know why, and reports to parents are becoming here and there more sane (sometimes against the opposition of the parents!). There is increasing emphasis on 'understanding', causes for the lack of which lie in the discrepancies between individual and institutional value systems.

In the guidance movement, which has so recently burgeoned from the mystic marriage of the bench and the couch, values have become quite explicit, particularly in the economic and political or power categories. Guidance has perforce been more individualized than much in our mass educational system, and specific conditions of person, time, and place have directed attention to individual needs for certain occupational and other rôles. Social demands and sanctions have been interiorized, and preferred behavior directively or non-directively indicated. At least to the extent that these things occur as a result of the guidance program, we can say that values have been intelligently handled, though perhaps largely on a utilitarian basis.

Evaluation is a word that is used frequently and easily by teachers and others without their realizing its relation to axiological problems. I have even heard teachers talk about evaluating a pupil, and without even batting an eye, and projects, even whole school programs, are often evaluated with the greatest of ease. There is no doubt an advantage in asking such questions as "Did you like it? Was it too easy or too difficult? Was the time profitably spent?" and so on. One must start somewhere. A professor once asked the director of a measurement bureau to help him make out an examination for one of his courses. The director asked him, "What are the objectives of the course?" This question so baffled the professor that he became quite angry, and indicated in no uncertain terms what he thought of people who would ask such a stupid question. In evaluation if anywhere values must be made explicit. They may be arrived at in various describable ways and judged on the basis of the means necessary for their attainment. The subgoals viewed as means to attain these values may be critically examined, and the methods employed in the attainment of the subgoals can be studied. Along the line it should be possible to do some measuring in order to determine the degree of effectiveness of the means and subgoals employed, or the extent to which the goal values were obtained and other values lost, or enhanced, in the process. This oversimplification perhaps suggests something of the complexity of the task of evaluation, of which many higher-level evaluators are quite aware. But they would be the first to admit, in the words of almost any doctoral thesis, that "further research is needed."

POSTULATES CONCERNING VALUE FOR EDUCATIONAL PSYCHOLOGY

If values are to have a more prominent place than they now hold in educational psychology, it should be possible to formulate a few tentative postulates which will serve to indicate something of what is involved, and also to summarize the point of view here presented:

1) Values, however named and categorized, are concepts, i.e., verbal constructs representing classes of objects or conditions cathected by individuals or groups.

2) While values as experienced are 'private' in the sense that they are a part of the inner life of the individual, they have significance only in the form of verbal or other behavior, in which form they may be studied directly and inferentially.

3) Value judgments are propositions about values that can be validated just as other propositions are validated, and hence properly come within the purview of the scientific method.

4) When value judgments are implicit, and they and the individual and social norms from which they are derived are not made explicit, semantic and other confusions in communication are likely to follow.

5) Value concepts may be of low or high generality, but whether values are viewed as means (individual or immediate goals) or as ends (more nearly universal or ultimate goals), evaluative judgments are based on known or anticipated consequences, or both.

6) Value systems (i.e., preference patterns) beyond such elementary physiological satisfiers as may exist, are not innate or fixed, but are acquired through a process of learning.

7) Learning within the institution of the school (and outside as well) involves modifications of values and value patterns, and such preferential behavior is therefore within the proper area of direct investigation of social scientists in general and educational psychologists in particular.

8) Since the values acquired by the individual act as determining tendencies which influence his behavior in ways that are of greatest importance to his own happiness and that of others, it follows that a significant part of the task of the educational psychologist is to contribute directly or indirectly to the value orientation of the members of the school community.

REFERENCES

1) Franz Alexander, "Values and science," *Journal of Social Issues*, vi, 4, (1950), 28–32.

2) G. W. Allport, and P. E. Vernon, "A study for personal values," *Journal of Abnormal and Social Psychology*, 26; 231–248, 1931, and *A Study of Values. Manual of Directions*, revised ed. Boston: Houghton Mifflin, 1931.

3) A. Bavelas, "A method for investigating individual and group ideology," *Sociometry*, 5 (1942), 371–377.

4) Department of Elementary School Principals, *Human Values in the Elementrary School*, Wm. Clark Trow, Warren A. Ketcham, and Others. Washington: National Education Association, 1952.

5) John Dewey, "Theory of Valuation," *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press, 1939, vol. 2, 67 p. Also, see Lepley, *op. cit.*

6) Educational Policies Commission, *Moral and Spiritual Values in the Public Schools*. Washington: National Education Association, 1951.

7) Herbert Fiegel, "The difference between knowledge and evaulation," *Journal of Social Issues*, VI, 4 (1950), 39–44.

8) George Geiger, and Others, "Values and social sciences," *Journal of Social Issues*, VI 4 (1950), 1–79.

9) C. L. Hull, "Moral Values, Behaviorism, and World Crisis," *Transactions of the New York Academy of Science*, 7 (1945), 90–94.

10) Max Hutt and Daniel R. Miller, "Value interiorization and democratic education," *Journal of Social Issues*, V, 4 (1949), 31–43.

11) Clyde Kluckhohn, *Mirror for Man*. New York: Whittlesy House, 1949.

12) Clyde Kluckhohn and Others, "Values and Value Orientations in the Theory of Action." In Talcott Parsons and Edward A. Shils (Eds.), *Toward a General Theory of Action*. Cambridge: Harvard University Press, 388–433.

13) Wolfgang Köhler, *The Place of Value in a World of Facts*. New York: Liveright Publishing Corp., 1938.

14) Roland Leipholz, *The Visual Arts in the Elementary School*. Unpublished doctoral dissertation, University of Michigan, 1953.

15) Ray Lepley (Ed.), *Value: A coöperative Inquiry*. New York: Columbia University Press, 1949.

16) Clarence I. Lewis, *An Analysis of Knowledge and Valuation*. LaSalle: Open Court Publishing Co., 1946.

17) Ward E. Madden, *Religious Values in Education*. New York: Harper and Brothers, 1951.

18) J. B. Maller and E. M. Glasser, *Interest Values Inventory*. New York: Bureau of Publications, Teachers College, Columbia University, 1939, 1943.

19) Daniel R. Miller and Max Hutt, "Value interiorization and personality development," *Journal of Social Issues*, V, 4 (1949), 2–30.

20) Charles Morris, *Signs, Language, and Behavior*. New York: Prentice-Hall, 1946.

21) Henry A. Murray, "Towards a Classification of Interactions," in Parsons and Shils (op. cit.) p. 463, also, Clyde Kluckhohn, H. A. Murray and D. M. Schneider, *Personality in Nature, Society, and Culture*. New York: Alfred A. Knopf, 1953, p. 23.

22) National Education Association, Commission on the Reorganization of Secondary Education, *Cardinal Principles of Secondary Education*. Washington: Department of the Interior, Bureau of Education, Government Printing Office, No. 34, 1918.

23) Talcott Parsons and Edward A. Shils (Eds.), *Toward a General Theory of Action*. Cambridge: Harvard University Press, 1952.

24) Eduard Spranger, *Types of Men: The Psychology and Ethics of Personality*. Halle: Max Niemeyer Verlag, 1928.

25) Symposium, in *Personality*. (April 1950), 1–74.

26) E. L. Thorndike, *Selected Writings from a Connectionist's Psychology*. New York: Appleton-Century-Crofts, Inc. 1949, Chapter XXI.

27) Wm. Clark Trow, "How shall we teach ethics?" *School Review*, 59 (December 1951), 519–524.

28) ——, "The public school a scapegoat?", The University of Michigan School of Education *Bulletin*, 23 (Novemeber 1951), 17–22; "Professional education and the disciplines: an open letter to Professor Bestor," *Scientific Monthly*, 76 (March 1953), 149–152.

29) ——, "A valuistic approach to religious education," *Religious Education*, 43 (May-June 1948), 169–174.

30) Wm. Clark Trow, Rosalind Zapf, and H. C. McKown, *Junior Citizen Series* (six booklets). New York: McGraw-Hill Book Co., Inc. 1951.

31) R. K. White. "Value analysis: a quantitative method for describing qualitative data, "*Journal of Social Psychology*, 19 (1944), 351–358.

32) Asahel D. Woodruff. "Personal values and the direction of behavior," *School Review*, 50 (1942), 32–42; "Personal values and religious backgrounds," *Journal of Social Psychology*, 22 (1945), 141–147; with Francis J. DiVesta "The relationship between values, concepts and attitudes," *Educational and Psychological Measurement*, 8 (Winter 1948), 645–659.

GROUP INTEGRATION IN A CASE DISCUSSION COURSE[1]

LEONARD A. OSTLUND

Oklahoma Agricultural and Mechanical College

I. THE PROBLEM

The purpose was to study group integration. Group integration as used herein concerns the nature of the structure of the whole, as it may be represented by concepts of field psychology. This encompasses the relationships between the various parts—students, teacher, physical and psychological components of the classroom, and the college environment—that make up the whole which is the psychological classroom situation as seen from the viewpoint of the students.

A limited aspect of the interpersonal relationships was surveyed. This facet was determined by observing the functioning of the individuals within the group. Functioning was operationally defined as the way in which the individuals behaved toward one another. The rationale underlying the many measurements and specific hypotheses to be discussed may be conceptualized in terms of the following fourfold theorem, given in its hierarchical order:

a) The greater the degree of positive affectivity between members of the group,
b) the greater the degree of group integration,
c) the greater the adequacy of group function,
d) the greater the productivity of the group.

The present study was concerned with probing the relationship between 'a' and 'b'.

II. IMPORTANCE OF GROUP INTEGRATION

Few will deny the value of group integration in the classroom. From the viewpoint of the individual personality certain individual needs can find their satisfaction in only group membership. From the viewpoint of the group, unless it satisfies individual needs, it will not long survive. From the viewpoint of the teacher, integration may further the learning process. From the viewpoint of social psychology, group integration denotes an equilibrium at an advanced developmental level, with unitary function as its salient characteristic. Group integration thus satisfies personal, educational, and social needs. A classroom group in rapport and bound together in the pursuit of a common goal is potentially more productive than is a less well-integrated group. The integrated group is superior because their social structure functions to make the resources of each member of the group available. The complexity of contemporary interpersonal relationships in the social field, as well as pressing national and international problems, make it mandatory for educators to provide students with concrete experiences in group problem-solving situations.

---

[1] This study was carried out under a Fellowship in the Department of Human Relations at the University of Kansas. The writer is indebted to the staff of the Department of Human Relations and to the staff of the Department of Psychology for constant stimulation; to Dr. Hilden Gibson, Chairman of the Department of Human Relations for his aid; to Dr. Edward G. Nelson for permitting the use of his class for the experiment; to Dr. Anthony J. Smith for statistical advice; and to his wife for constant assistance.

III. THE SUBJECTS AND THE SETTING

The subjects were twenty-five male college students in a case-discussion course at the University of Kansas. Twenty were juniors, five were seniors. They were studied in the spring semester of a one-year case-discussion course, therefore, some students were acquainted from the previous semester and all had experienced one semester in this type of case-discussion course.

In this classroom, students and teacher are seated around a large table, vis-a-vis. The teacher plays a permissive rôle, calculated to enhance spontaneity of expression by the students. Barker and Wright (*1*) have suggested the importance of the setting in which interpersonal relationships take place. Behavior settings are coercive in that they tend to elicit activity appropriate to them. The writer found this a valuable guiding principle.

IV. METHODOLOGY AND RATIONALE

It was considered fruitful to study group integration from the viewpoint of the members of the class, to assess the affective aspect of the psychological environment by determining their feelings toward one another. Experience suggested that one useful index of affectivity would be the choice of associates, and that the pattern of choices could be related to group structure. The choice pattern was determined by sociometric techniques as outlined by Moreno

(*10*). These were used to portray interpersonal classroom relationships. Pioneer work in this area has been done by Berrien (*4*), whose book, *Comments and Cases on Human Relations*, contains tentative research findings. An enduring interest in this area was engendered by a stimulating year with Dr. Berrien at Colgate University as colleague in teaching the Human Relations Course.

When rapport had been established, the following questions from which the sociometric data have been collected were read to the class:

1) With what five members of this class do you now spend most of your leisure time?
2) With what five members of this class, would you—if you could—spend most of your leisure time?
3) Let us assume a problem comes up in class on which you can work in small groups. Choose the five with whom you would most like to work.

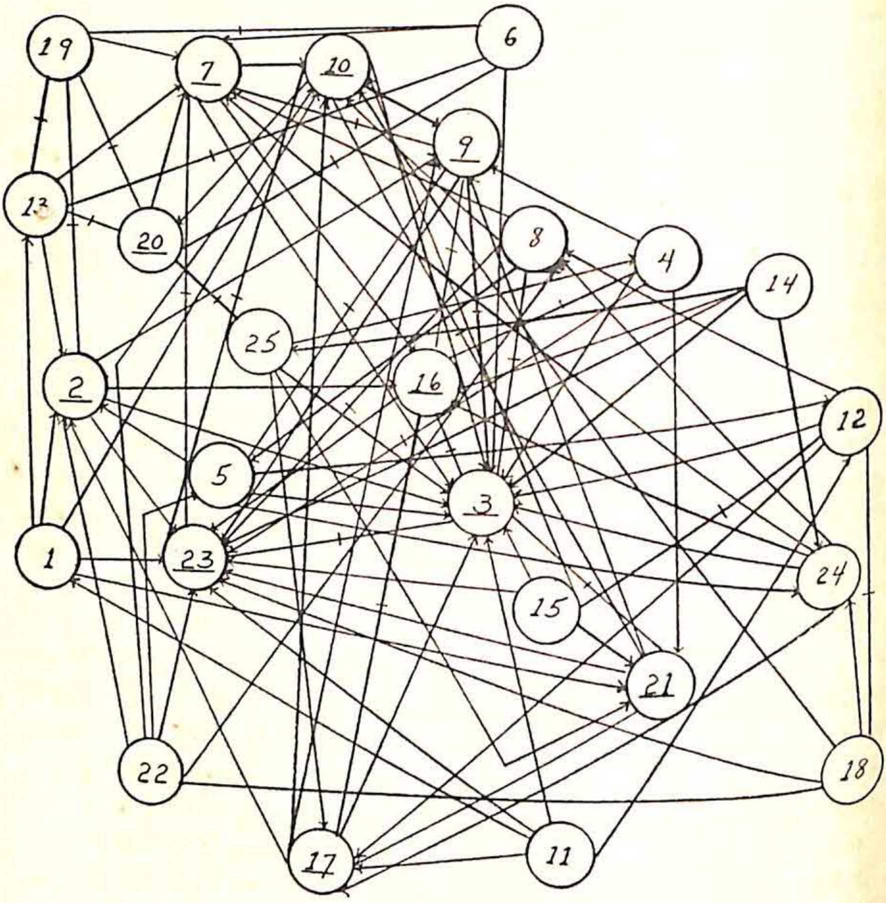
As mentioned previously, group integration, tested sociometrically, was approached primarily from the viewpoint of the students. However, it is apparent to any educator that the group does not function independently of the teacher and first-hand observation suggests that though the teacher may be physically absent, he is nevertheless psychologically present. The writer has previously affirmed his conviction concerning the importance of the atmosphere created by the teacher and the setting. Therefore, the writer felt it would be productive to explore the teacher's knowledge of the group's structure.

The Bard of Avon has sagely observed, "It is a wise father that knows his own child." One might say, "It is a wise teacher that knows his own class," because of the gulf in years, knowledge, and ideals which sometimes separates teachers from students. This aspect was probed by asking the teacher for three rankings as follows:

1) Predict the order in which the students will be chosen as work associates.
2) Predict the order in which the students will be chosen as leisure-time associates.
3) Rank the students according to grade standing.

V. DATA AND INTERPRETATION

A) *Sociometric analysis.*—A work sociogram was drawn up from the data of question 3). It reveals a complex well-integrated social



WORK SOCIOGRAM

Key:

Mutual Choice
(A chooses B)
(B chooses A)
A —+— B

One-way Choice
(A chooses B)
(B does not choose A)
A ——→ B

Underlined figures -- Five or more choices received.

structure. Many people are chosen by others, several individuals receive many choices. Further evidence of group integration is the appearance of well-integrated subgroups, denoted by mutual choices between three and four individuals, yet none of these subgroups are removed from the group as a whole. Moreover, the whole is connected by many chains of relationships whose peripherations are far-reaching. The sociogram points to a hierarchical differentiation as follows:

1) Leaders—those who receive many choices.
2) Followers—those who receive an intermediate number of choices.
3) Isolates—those who receive no choices.

Our understanding of social structure may be augmented by applying the concepts of other social scientists. Barnard (*2*) has emphasized the distinction between scalar and lateral lines of organizational structure. The term scalar refers to vertical lines which go up-to and down-from the organizational hierarchy. The term lateral refers to lines which go between members on the same level, such as worker-worker, etc. There is general agreement that effective organizational structures are articulated by an interlocking network of both lateral and scalar lines. The sociogram manifests these characteristics. Many lines go to those who may be characterized as being on the 'top level', in terms of the number of choices; for example, numbers 3 (eighteen lines or choices), and 23 (fifteen lines or choices). There are also many interconnections on the lower levels.

Another contribution to the understanding of group structure has been made by Bavelas who has extended Lewin's topology by techniques that permit a mathematical analysis of group integration in terms of hodological space. Bavelas (*3*) has suggested that rapidity, direction, and extent of spread of communication depend partly upon the pattern of connections between individuals in a group. In this frame of reference, we may consider each individual as a center of communication and the lines between individuals as links of communication. Applying this hypothesis to the sociogram, it seems reasonable to assume that, within such a well-integrated group, communication should spread rapidly and become widely disseminated.

While the preceding techniques permit us to represent the structure of the group, other techniques must be utilized to fathom the causal properties of these interpersonal relations. Observations of the individuals functioning in this group and in other groups, as well as socio-economic data, may provide valuable clues. Because the primary purpose of this research was an attempt to delineate group structure rather than to explain it, only a brief example will be given. One member who received many choices had a high scholastic standing. His participation was high in frequency and quality, and he evinced a general skill in communication. The reason for his high number of choices was apparent. However, another who received many choices was virtually a non-participant in class-discussion. The reader would be correct in postulating some prestige factor unrelated to the classroom situation. Though a mediocre student, he was a well-known athlete. These examples of qualitative and quantitative analysis illustrate rich possibilities in exploring the factors which determine group structure.

B) Statistical validation of sociometric data.—(1) Introduction: The sociogram was constructed from one variable—the students' choice of work associates. Two students were termed isolates—they received no choices. However, to gain a more adequate understanding of an individual's acceptability—in terms of choices received—it is necessary to sample different kinds of criteria. It may happen that the student who is chosen for intellectual ability may not be chosen for athletic prowess. However, such a situation need not unnecessarily concern a teacher, for this individual may not feel psychological isolation as his abilities have been recognized by his peers in the intellectual area. Therefore, in the following statistical validation of the sociometric data, the writer chose to deal with combined variables, since the group endures, to some extent, beyond the confines of the classroom.

All three variables—present leisure-time associate, preferred leisure-time associate, and preferred work associate—were combined into a total choice pattern for each individual, yielding the pattern for the group as a whole. The sociometric indices that follow have have been developed by Bronfenbrenner (*5*). The reader will find a full development of the rationale and statistical computations in the reference given.

(2) The Sociometric Indices: In the Social Choice Index, Table 1, the writer has departed from Bronfenbrenner's terminology which considered the total number of choices received by a subject the 'Social Status Index' in favor of the more accurate, 'Social Choice Index' suggested by Loomis (*6*). The values computed were:

1) The Mean .... $\bar{X} = 15.$
2) The Standard Deviation .... $\sigma = 5.97$
3) The Appropriate Degree of Skewness .... $A_3 = .097$

The mean and the standard deviation permit the transformation of each score into a standard score equivalent, which can be checked in Salvosa's Tables (*11*) under the appropriate degree of skewness thus yielding the probability of chance occurrence.

Table 1.—Social Choice Index

I Student Number	II Number of Choices Received	III Upper Limit	IV Raw Score Deviation	V Standard Score Equivalent	VI P. of Chance Occurrence
	S	L	L-M	$\frac{L\text{-}M}{\sigma}$	P.
22	2	2.5	−12.5	−2.093	*.0157
1 11	3	3.5	−11.5	−1.926	*.0245
15 18	5	5.5	− 9.5	−1.591	.0527
14 4	7	7.5	− 7.5	−1.256	.1037
5	10	10.5	− 4.5	− .753	.2270
2 24	11	11.5	− 3.5	− .586	.3040
12	12	12.5	− 2.5	− .418	.3430
13	14	14.5	− .5	− .083	.4730
17 8	15	15.5	.5	.083	.5390
21 25	16	16.5	1.5	.251	.6047
16	18	18.5	3.5	.586	.7240
20 10	19	19.5	4.5	.753	.7765
6 9 19	23	23.5	7.5	1.256	.8935
7	27	27.5	12.5	2.093	*.9793
23	32	32.5	17.5	2.931	†*.9975
3	38	38.5	23.5	3.936	†*.9998

* P. < .05
† P. < .01

The Social Choice Index provides a clue to group integration. The scores of three students are so low that they could occur by chance only three times in a hundred. On the other end of the continuum, one student received so many choices that this could have happened by chance only three times in a hundred. He is surpassed by two others whose choice totals are so large that this could have happened by chance only once in a thousand times. Our statistical treatment permits us to state that such results were not due to adventitious circumstances but may be attributed to the social determinants in the classroom situation.

The number of isolates has been considered an indication of group integration. In the combined data, no isolates appeared. The probability of zero isolates as computed by the Index of Isolation was .0096 ($P. < .01$). This is a statistically significant indication of group integration.

The number of reciprocal choices is important to group integration, for this reveals the number of individuals who are in accord with one another. The total used is the total mutual choices for each category computed separately which amounted to eighty-nine. A computation of the Index of Coherence yielded a value of .00001 ($P. < .01$), which affords further indication of group integration. Thus all three indices have been found statistically significant and the writer interprets this as complementing the conclusions concerning group integration which were based on the work sociogram.

C) The application of correlation statistics.—(1) *The Phi* Coefficient of Correlation: It is important to determine whether the same individuals were chosen for two categories. The question explored may be phrased as follows: Does student #1 tend to choose student #2 as both preferred leisure-time associate and present leisure-time associate? The phi coefficient yielded a value of .583 ($P. < .01$). We may therefore assume that those preferred as leisure-time associates are the present leisure-time associates, that the students tend to actualize their preferences in this area. In computing the rank correlation coefficients which follow, the writer discarded the category, present leisure-time associate, in favor of preferred leisure-time associate, so that both work and leisure-time choices would be an expression of preference. Table 2 presents a list of all the correlations.

(2) *The Spearman rank difference correlations:* The writer has accepted the sociometric data as supporting the hypothesis of group integration. This hypothesis was next tested in a different way. The writer believed that under conditions of group integration, an individual's score in one category would be positively related to the score that this individual would receive in the other. This would be consistent with Homans' (7) hypothesis that the feelings generated within one system of relationships (the classroom) would tend to perpetuate themselves by flowing into another system (the extra-curricular and social). Then, too, it must be remembered that the specific task for which associates were chosen was that of working on a classroom problem in a small group. This would in-

TABLE 2.—CORRELATION COEFFICIENTS*

	C.C.	Sig. Level
A. *Spearman Rank Difference Correlations:*		
1) Teacher's prediction of work associate and student's choice of same	.617	P. .01
2) Teacher's prediction of leisure-time associate and student's choice of same	.529	P. .01
3) Teacher's prediction of work-associate and student's grade standing	.424	P. .05
4) Teacher's prediction of leisure-time associate and student's grade standing	.193	Not sig.
5) Teacher's prediction of leisure-time associate and teacher's prediction of work associate	.583	P. .01
6) Student's choice of work associate and student's choice of leisure-time associate	.750	P. .01
7) Student's choice of work associate and student's grade standing	.490	P. .05
8) Student's choice of leisure-time associate and student's grade standing	.270	Not sig.
B. *Phi Coefficient of Correlation:*		
1) Student's choice of leisure-time associate and student's present leisure-time associate	.583	P. .01

* The writer has recognized a statistical problem in that it is, to some extent, a matter of judgment upon the part of an experimenter whether the teacher may be considered as a criterion. This is not an academic problem for on this decision rests the appropriateness of the correlation formula to be used. In order to resolve this problem and in the interests of experimental rigor, the writer has calculated these eight coefficients by formula A and B. (*8*) The differences ranged from .015 to .05, but in no case would the use of the other formula have changed the statistical significance of the results. The formulas used are those suggested as appropriate by Kendall for data of this nature.

volve intimate interpersonal relations common to both systems of activity. The coefficient between the students' choice of leisure-time associate and the students' choice of work associate was .750 ($P. < .01$). The writer interprets this result as further support for his hypothesis of group integration.

The writer did not hypothesize concerning the relationship between the students' choice of work associate and the students' grade standing. On the one hand grades are based on the ability to analyze a case in a written report, as well as the ability to communicate and participate in the class discussion. One could therefore predict a positive relationship because of the similarity between the grade criteria and the task. A positive relationship might also be argued on the hypothesis that, since students know one another's grade standing, they would tend to work with the best students in order to gain good grades. On the other hand, it might be argued that while work and leisure-time associates are highly correlated (P. $< .01$), this may be due to a global interpersonal attraction, but that this is not necessarily contingent upon a student's grade standing. Then, too, social pressures might influence the individual to choose as a work companion, a failing student who was a fraternity brother or friend. The Spearman Rank Difference Correlation Coefficient was .490 (P. $< .05$). Thus there is a factor common to both, but the data do not allow us to state its nature. The writer interprets this finding as favorable to integration in that, since it is positive, we do not have a social climate in which a student's grade standing works against him and has the effect of his not being chosen as a work associate.

Concerning the relationship between a student's grade standing and his desirability as a leisure-time associate, the writer hypothesized that these would not be related, since grade standing is seldom a determinant of social acceptability. The Spearman Rank Difference Correlation Coefficient was .270 which is not significant—one of the lowest correlations reported in this paper.

In the present experiment, the primary concern with the sociometric data was experimental rather than pedagogical. Nevertheless, the practical value was not overlooked. After the teacher had ranked the students the writer divulged the sociometric data, which were then used by the teacher in furthering the goals of the course.

The correlation between the student's choice of work associate and the teacher's prediction of the student's choice of work associate was .617 (P. $< .01$). The correlation between the student's choice of leisure-time associate and the teacher's prediction of the student's choice in this area yielded a coefficient of .529 (P. $< .01$). Thus in both areas, the teacher was able to predict the students' choices to a degree inexplicable by chance. The writer believes the methodology is one factor which enables the teacher to know the structure of the group.

With respect to the relationship between the teacher's prediction of the students' work associates and the students' grade standing, the coefficient was .424 (P. $<$ .05). We may, therefore, state that the teacher tended to predict that students who had the highest grades would receive the highest number of choices as work associates. The relationship between the teacher's prediction of the student's choice of leisure-time associate and the student's grade standing was also computed. The coefficient was .193, statistically insignificant, the lowest correlation reported in this study. Since it was found that there was also little relationship between the students' choice of leisure-time associates and the students' grade standing, we may conclude that, from both the viewpoint of the teacher and the students, grades have little relationship to choice of leisure-time associates.

The relationship between the teacher's prediction of work associate and his prediction of leisure-time associate was positive and significant, the coefficient was .583 (P. $<$ .01), and mirrored the relationship between the students' choices in these areas.

In conclusion, the teacher's predictions accurately reflect the choices made by the students. The writer interprets this as important to group integration in that the teacher evinces a sound understanding of his class' social structure. When the teacher is able to perceive the social situation as experienced by the students, an attitude which Mead has described as assuming the rôle of the generalized other (*9*), this common perception should lead to greater clarity of communication between the teacher and the students which, in a permissive atmosphere, may be expected to facilitate group integration.

VI. SUMMARY AND CONCLUSIONS

The results suggest that the hypothesis of group integration is tenable. Sociometric analysis provided qualitative indications, and the sociometric indices were found to be statistically significant. Correlation coefficients affirm that the relationships between student-student variables and teacher-student variables cannot be attributed to chance. However, the writer cautions against generalizing these findings to other case-discussion courses. Any group that is together over a period of time develops some kind of organization. Further research is needed concerning the degree of integration within other courses, as well as exploration of other variables of group organization.

From the viewpoint of the writer, the causal nexus is that these students shared common goals while interacting as a group, in a permissive environment. Individual needs could be satisfied only through group problem-solving; conversely, the group could function adequately only if individual needs were satisfied. This manifests the interdependence of group and individual activities. Individual and group satisfaction are complementary. When individuals find personal satisfaction and group acceptance in their interpersonal relationships, group integration emerges.

REFERENCES

1) Roger G. Barker and Herbert F. Wright. *Methods in Psychological Ecology*, Lawrence: University of Kansas, 1950.

2) Chester I. Barnard. *The Functions of the Executive*, Cambridge: Harvard University Press, 1947.

3) A. Bavelas. "A mathematical model for group structure," *Applied Anthropology*, Vol. 7, No. 3, Summer, 1948, pp. 16–30.

4) F. K. Berrien. *Comments and Cases on Human Relations*, New York: Harper & Bros., 1951.

5) Urie Bronfenbrenner. "A constant frame of reference for sociometric research," *Sociometry*, Vol. 6, No. 4, November, 1943, pp. 363–395.

6) C. P. Loomis and C. H. Proctor in *Research Methods in Social Relations*, M. Jahoda, M. Deutsch, and S. W. Cook, Eds., New York: The Dryden Press, 1951.

7) G. C. Homans. *The Human Group*, New York: Harcourt, Brace & Co., 1950.

8) Maurice G. Kendall. *Rank Correlation Methods*, London: Charles Griffin & Co., Ltd., 1948.

9) G. H. Mead. *Mind, Self, and Society*, Chicago: Chicago University Press, 1934.

10) Jacob L. Moreno. *Who Shall Survive?*, Washington, D. C.: Nervous and Mental Disease Publishing Co., 1934.

11) L. R. Salvosa. "Areas, ordinates, and derivatives of the Pearson Type III Curve," *Annals of Mathematical Statistics*, Vol. 1, No. 1, February, 1930, Appendix, pp. 1–187.

SOME RELATIONSHIPS BETWEEN STUDENTS' PERCEPTIONS OF SCHOOL AND THEIR ACHIEVEMENT

LESLIE F. MALPASS

Psychology Department

Southern Illinois University

Many studies have been made which relate the intelligence, interests, aptitudes and personality characteristics of students to their achievement in school. Relatively few experiments have been reported which indicate how the school appears to the students and how this factor might be related to their scholastic success. The purpose of this study was to investigate some relationships between students' perceptions of various aspects of the school situation and selected criteria of school achievement. The general hypothesis of the study was: *There is a relationship between measures of students' perceptions of school and achievement in school.*

METHOD

Sample of Students.—Ninety-two students, representing almost the entire eighth-grade population of the Solvay, N. Y., public school system, were selected as the student sample.[1] This school system was selected because students enrolled came from family backgrounds which were proportionately representative of socioeconomic stratification in upper-New York State. All students had been subjected to the same teachers and general school influences for at least two years; the majority had been enrolled in the same grammar and junior high schools for at least five years. In this way, some control over the effects of school factors on students' achievement was obtained.

Measures of Perception.—All subjects were tested with three instruments designed to measure perceptions of school; a sentence completion test consisting of fifty incomplete sentences; a ten-card apperception test, the pictures of which depicted children in various school settings; and an autobiographical-type composition in which students expressed their feelings about various aspects of school.

---

[1] Mr. John McAnaney, Guidance Director of the Solvay Schools, and Miss Anna Murtaugh, Principal of Intermediate School, were extremely coöperative, and the writer is indebted to them and the teachers involved.

Each of these instruments was subdivided into five areas in an attempt to obtain indices of students' perceptions of their teachers, classmates, achievement in school, school discipline and school in general. Thus, total test scores and five subarea scores for each of the three perceptual tests were available for correlation with the achievement criteria.

A five-point rating scale was devised in order that responses to the various perceptual test stimuli could be judged by independent raters:

> Rating 1 indicated that the response portrayed a definitely positive perception of the variable in question (pleasure about, liking for, acceptance of in a positive manner, or hopeful about).
>
> Rating 2 indicated that the response portrayed probable positive feelings about the aspect of school in question.
>
> Rating 3 indicated that the response portrayed neutral or otherwise unscorable responses (purely descriptive, no feelings expressed, or insufficient material to rate in any other way).
>
> Rating 4 indicated that the response portrayed probable negative feelings about school (displeasure about, dislike of, anger or hostility towards).
>
> Rating 5 indicated that the response portrayed definitely negative feelings about the aspect of school in question.

In addition, each response was rated with respect to the subarea of school towards which the response was directed. This was necessary because, for example, although on the Personal Document Test a student might be asked to write about a particular aspect of school, e.g. "My Teachers," his response might also include some statements about how he felt about his school work, discipline, and so forth.

A prestudy group of fourteen eighth-grade students (seven boys and seven girls, with intelligence quotients within the normal range) was used to determine whether the tests would elicit unique, spontaneous expressions from the students without their becoming aware of the "kind of inference the experimenter intended to make" (*10*, p. 215). Results from this group, evaluated by three judges independently, confirmed the validity of this assumption.

DESCRIPTION OF THE PERCEPTUAL TESTS

A) *The Sentence Completion Test* (SCT).—In its original form, seventy-five items were selected for this test but twenty-one were eliminated by a board of three psychologists on the basis of being too ambiguous, too highly structured so they elicited only one type of response, or too difficult in grammatical construction to yield other than 'neutral' replies. Four additional questions were discarded to round off the number of items at fifty, leaving ten incomplete sentences for each of the five school areas tested. The following pattern was set up for administration purposes:

Sentence Number	*School Area*
1, 6, 11 . . . . 41, 46	School in general
2, 7, 12 . . . . 42, 47	Teachers
3, 8, 13 . . . . 43, 48	Classmates
4, 9, 14 . . . . 44, 49	Discipline
5, 10, 15 . . . . 45, 50	Achievement

The test was administered in group form and results were analyzed by use of the five-point rating scale described above. A description of this test has already been reported (*6*).

B) *School Pictures Test* (SPT).—This test was devised and has been used by a group of advanced graduate students at Syracuse University, under the direction of Arthur W. Combs. In its original form, the test contained twenty pictures. However, analysis of the stories given by the prestudy group showed that there tended to be much repetition on several of the cards and that, in fact, practically as valid an interpretation of feelings could be made from ten cards as from all twenty. Therefore, ten cards, two representative of each of the five school areas under consideration, were selected for the final experiment.

In its construction, purpose, use and administration, the School Pictures Test is similar to other apperception tests, with the exception that all pictures are structured in the school area only. However, scoring procedures for this experiment were slightly different from those commonly used; the five-point rating scale already described was used to quantify the responses received from SPT. A complete description of this test is available elsewhere (*6*).

C) *Personal Document Test* (PDT).—In addition to obtaining data from the sentence completion and apperception tests, it was felt advisable to administer a test which would provide spontaneous, verbatim statements from the students themselves about the various aspects of school in question. This test was divided into five areas, entitled, "My School," "My Teachers," "My Classmates," "My Work in School," and "Behavior in School." The test was presented to the subjects in a group (studyhall) situation, using standardized directions for writing the composition. Two full periods for completion were permitted by school authorities where necessary. Most students spent from thirty-five to fifty minutes on the task.

The PDT was analyzed and scored on the basis of the same five-point rating scale used with the other tests. Like them, also, judgments were made as to the aspect of school represented in each response since, in spite of topical headings outlined for each section, more than one feeling-tone was expressed in all the compositions. Consequently, each student tended to yield a different number of rated statements about each school area. In order to obtain a subarea score, then, all responses relating to that area were summed and the mean score obtained was used as the subarea score.

Reliability of ratings.—Four judges, including the examiner, independently analyzed and scored fifty SCT protocols, thirty SPT and thirty PDT protocols. Inter-judge correlations ranged from .78 to .92 for the three instruments. In addition, re-test reliability estimates, based on re-rating of these protocols six months later by the examiner, were .88 (SCT), .80 (SPT) and .89 (PDT).

It was felt these reliability estimates were high enough to justify the assumption that experts tended to rate responses to the three perceptual tests essentially the same and that the examiner showed consistency in his ratings.

When the same four judges independently indicated the school areas to which individual responses belonged, percentage agreements ranged from seventy-four to one hundred with the examiners' ratings. Again, it was felt the agreement was high enough to assume that experts tended to agree with the examiner in their judgments of the aspect of school indicated by students' responses.

A correlation between the SCT and SPT of .71, suggests that a large common element was being measured by these two instruments. The correlation between the SCT and PDT was not so high (.58) and the correlation between the SPT and PDT was .52. The lower correlations between the PDT and the other tests may be a function of the PDT being a less-disguised projective method than either the SCT or SPT. It should be pointed out that there were no empirical validity estimates available for any of the three perceptual tests; justification for their use was based on pre-experimental group experience with them and the logical validity which is commonly accepted for projective instruments.

Achievement criteria.—Two measures of achievement, end-of-semester grades and standardized test scores, were used. It was not presumed that these two methods measured achievement in the same way. School grades are subjective (teacher) ratings of students' class work, while standardized achievement tests provide a more objective measure of how much a student knows about a particular subject matter, regardless of such factors as intelligence, personality traits, age, or ability to get along with others.

End-of-semester grades in four basic curriculum subjects—English, mathematics, social studies, and general science—were assigned number-values from one to five, corresponding to letter grades A through E (failure). The mean score for the four subjects was obtained, and this score was used to represent the grade-score for each student. Because of time limitation, only the Stanford Reading Test and Stanford Arithmetic Test were administered from the entire Stanford Achievement Test Battery. These yielded relatively objective measures of achievement.

Intelligence criterion.—In order to partial out the effects of intelligence on achievement, the California Test of Mental Maturity S-Form was used. The Total Test IQ was used as the index of intelligence. The mean IQ of the sampling group was 105.30, with a standard deviation of 12.51.

Statistical procedures.—Product-moment correlations were computed between total and subarea perceptual test scores and the three achievement criteria, using Shephard's correction because of coarseness of groupings in the perceptual test ratings. Partial correlation coefficients were then computed, thus mediating the effects of mental ability on achievement. The latter data were of primary concern to the findings of the study.

RESULTS

Table I represents the relationships between the measures of perception and achievement with the effects of intelligence partialed out. The findings may be summarized as follows:

1) The correlations between mean total scores of all three perceptual tests and mean end-of-semester grade scores were signifi-

TABLE I.—PARTIAL CORRELATION COEFFICIENTS

Test	Stanford Arithmetic	Stanford Reading	Mean Grades
Sentence Completion Test	.27	.06	.57
School Pictures Test	.13	.06	.45
Personal Document Test	.02	−.03	.31
Subareas of Perceptual Tests			
School in General			
Sentence Completion Test	.14	−.03	.39
School Pictures Test	.02	.07	.25
Personal Document Test	.23	.05	.14
Teachers			
Sentence Completion Test	.08	−.01	.57
School Pictures Test	.15	−.02	.39
Personal Document Test	.02	−.02	.08
Classmates			
Sentence Completion Test	.01	−.01	.29
School Pictures Test	.22	.07	.31
Personal Document	−.03	.04	.12
Achievement			
Sentence Completion Test	.30	.03	.48
School Pictures Test	−.04	.04	.31
Personal Document	.15	.02	.48
Discipline			
Sentence Completion Test	.26	−.02	.38
School Pictures Test	.22	.07	.32
Personal Document	.02	−.05	.16

cant at the one per cent level of confidence, although the correlation between the PDT and grades was quite low (.31).

2) All correlations between subarea scores on the SCT and SPT and end-of-semester grades were significant at the one per cent level of confidence but only one subarea score on the PDT (interestingly enough, the subarea representing students' comments about their own school work) was significant at this level.

3) Although approximately one-third of the correlations between total and subarea perceptual test scores and Stanford Arithmetic Test scores were significant at the five per cent level or better, they were all so low as to suggest little relationship between the variables.

4) There were no significant correlations observed between

measures of perception and the Stanford Reading Test. All correlations, both between total and subarea perceptual test scores and the reading test scores, clustered around zero, with almost as many negative as positive correlations. Obviously, no relationship occurs between these measures of perception and achievement.

Thus, the conditions of the hypothesis, *There is a relationship between measures of students' perceptions of school and their achievement in school*, were fulfilled with respect to one type of achievement criterion, end-of-semester grades, but were rejected with respect to the second type of achievement criterion, standardized achievement tests.

CONCLUSIONS AND DISCUSSION

On the basis of obtained results, the following conclusions seem appropriate: 1) Students' perceptions of school, and various aspects of school, seem to be related to achievement in school as measured by end-of-semester grades. It is not possible, on the basis of present evidence, to determine a cause-effect relationship between the two variables but speculation leads to one of two possibilities. First, suppose that the child's low grades are caused by negative feelings about school. Logically, one might expect the following cycle to take place: The child dislikes school, obtains low grades, dislikes school even more, and continues to get low grades. On the other hand, suppose that negative feelings about school are caused by low grades; the converse of the above process takes place, but with the same results accruing—low grades and dislike of school, or high grades and positive feelings about school.

It is significant that both the SCT and SPT subareas suggest significant relationships between grades and how students view such specifics as teachers, discipline, school work and peers, as well as the generalized concept of 'school'. Several studies in the past have shown relationships between attitudes towards teachers and grades (*4*, *5*) but none other has shown relationships between all these areas in one experiment.

2) Students' perceptions of school, and of various aspects of school, do not seem to be related to achievement as measured by standardized achievement tests. The weight of evidence suggests that there is little relationship between how children view their school situation, and certain aspects of it, and their objectively-measured knowledge of arithmetic and reading. These findings are

in general agreement with Sawin's findings in the arithmetic area (*8*) but disagree with those pointed out by Sheldon (*9*) in the reading area. This suggests that further work in this area is needed.

SUMMARY

Modern theory of perception holds that how a person views himself and his world governs his behavior. This study, which has been exploratory in nature, set out to ascertain some relationships between students' perceptions of school and their achievement in school, when mental ability was held constant. Results obtained were equivocal. Little or no relationship existed between measures of perception and standardized achievement test scores, which are objective measurements of the cumulative effects of knowledge in particular subject matter; significant relationships were found between measures of perception and end-of-semester grades, which are subjective evaluations of scholastic achievement predicated on the immediate effects of knowledge. In view of the latter correlations, further consideration of the ways in which students view their school situation provides a suitable and profitable area of educational research.

REFERENCES

1) Gordon W. Allport, "The use of personal documents in psychological science," Social Science Research Bulletin #49, 1942.

2) Oscar L. Buros. *Third Mental Measurements Yearbook*. New Brunswick: Rutgers University Press, 1949.

3) L. B. Eron, D. Terry, and R. Callahan. "The use of rating scales for emotional tone in TAT stories," *J. Consult. Psych.* XII (1940), 31–38.

4) H. H. Remmers. "Relationships between students' marks and attitudes towards instructors," *School and Society* XXVIII (1928), 759–60.

5) ——. "Further studies in attitudes: Series III" *Studies in Higher Education*. Purdue University, 1943.

6) Leslie F. Malpass. *Some Relationships between Students' Perceptions of School and Selected Criteria of Achievement*. Unpublished doctor's dissertation, Syracuse University, 1952.

7) Helen Sargent. Projective methods: their origins, theory and applications in personality research, *Psych. Bull.*, XLII (1945), 257–93.

8) Enoch I. Sawin. *Motivation in Mathematics*. Unpublished doctor's dissertation, University of Chicago, 1951.

9) William D. Sheldon. *A Study of College Students with Scholastic Difficulties*. Unpublished doctor's dissertation, Syracuse University, 1948.

10) Robert White. "Interpretation of Imagination Productions." *Personality and the Behavior Disorders*, ed. J. McV. Hunt. New York: The Ronald Press, 1944.

COMMUNICATION THEORY AS THE UNIFYING THEME IN TEACHING EDUCATIONAL PSYCHOLOGY

SAMUEL Z. KLAUSNER

The Hebrew University, Jerusalem, Israel

Those who have taught the beginning course in educational psychology have recognized the difficulty in unifying the diverse material on motivation, measurement, child development and learning and, at the same time, presenting it in such a way that it will aid the prospective teacher in understanding the child and in guiding him in his life adjustments. Efforts to organize the course vary from a formal systematic approach to plans centering in one or another aspect of the material which the author or teacher considers to be of fundamental importance. It has, perhaps, been difficult to order the materials in a developmental sequence from basic definitions to final generalizations since educational psychology is not a discipline in the formal sense but an applied field drawing on many disciplines.

One attempt to achieve integration has been made by those who would organize the course around child study. This method unifies the material by bringing it to bear in understanding the growth and development of the child as a whole. This may be said to represent the individual psychological approach. Another conceptual frame for organizing educational psychology might be called the social-psychological. Those who organize the course around personality in society and culture would fall into this category.

Still another example of the social-psychological approach will be discussed in this paper. This is a description of an attempt to use communication theory as a central theme in the development of the course. It is suggested that this approach may accomplish two objectives: it serves to integrate the seemingly scattered materials which are drawn from several social sciences, and it brings the individual behaving as a whole into clear focus within a complete social-cultural context. This comes about by viewing the manner in which society, culture and personality influence the meanings which an individual acquires and which form a basis for personal and social action.[1]

The course opens with a view of the classroom as a system of interacting individuals. Each person interacts with the others by communicating symbols. These symbols are made up of signs, which are the objective facts of the environment such as sounds, sights or movements, plus the meanings which the respective individuals give to these signs. Pupil and teacher behavior are, in good measure, functions of the symbols which they perceive in their environment.

A principal objective is to help the prospective teacher to an awareness that the meanings which the learner is constantly perceiving may be quite different from those that the teacher may intend to communicate. We explore the way the same objective facts of the environment, whether they be auditory, visual, kinesthetic or other stimuli which we have referred to as signs, may be given different meanings by different pupils depending on their personalities and the social and cultural background from which they view them. The course revolves about the manner in which symbols emerge or evolve for an individual as he attributes his meanings to the signs in the environment.

To illustrate: we discuss how a simple critique by a teacher of a pupil's statement may be understood by one pupil as insinuating his stupidity or by another as a vindication of his intelligence. Some pupils may even find in it a suggestion of racial discrimination or yet another pupil may see it as suggesting a better way to attack the problem at hand.

Since a pupil's behavior is partially determined by the symbols which he personally develops, or by the meanings which he reads into the signs in his environment, it becomes important for us to study some of the influences which determine the way in which a sign gets its meaning and so becomes a symbol. Thus, we try to understand such things as his social and cultural background and his attitudes towards his environment and himself. All of these enter into the context from which the meanings develop. This also implies discussion of perceptual processes and the way in which they structure the meanings which an individual acquires.

---

[1] This description is based on a one semester course which was given at The City College of New York. It has lent itself to adaptation to new cultural conditions and is now being taught as a year course at The Hebrew University in Jerusalem, Israel. An outline of the course of study is available upon request from the author.

For example, we draw upon sociology to understand the social context in which meanings emerge. Conditions of social stratification are considered as they may influence, for instance, the different ways in which working-class and middle-class pupils may interpret the communication of a middle-class teacher. Thus, we are interested in what happens when a teacher condemns a boy for physically striking back at an assailant after his father has told him that any other behavior in that situation would be unmanly.

Sociology is further applied in studying social rôles. We investigate the manner in which individuals use symbols to tell others what behavior they expect in certain situations. As an instance of this, we would be interested in the meaning which a pupil might see in the teacher's leaving the room if he were involved in an authoritarian teacher-pupil rôle. How would the meaning differ for a pupil interacting in a democratic situation? What problem would arise for the pupil who had been accustomed to the authoritarian and then finds himself in a class where democratic rôles are the rule? Can the teacher do anything to help change the meanings which this pupil might impute to the situation?

Anthropological findings help us to understand the influence of the cultural ethos in producing meanings. For instance, the problem that the teacher faces in communicating with pupils from varying cultural backgrounds is discussed. How do they implicitly judge the teacher's behavior when it does not correspond to the folkways and the mores which they have learned to accept as proper? Thus, the class might study the problems that the northern-bred teacher would face in teaching in a southern school.

Studies in personality enter as we talk of the development of the self and the way in which individuals refer their meanings to it. Does learning to read appear as a threat to be avoided simply because an individual has learned to think of himself as inadequate?

Research in developmental psychology does not come under consideration as a dry report of facts and figures. Rather, we ask about the effect of particular developmental stages or the effect of being a short boy among tall girls upon the meanings which an individual reads into his environment.

Then, we are prepared to understand the problem of communicating academic or subject matter meanings. The prospective teacher can now see that teaching linear equations is not simply a question of logical step by step presentation with adequate provision for discussion and review, but also a problem in dealing with a personal-social-cultural matrix in which the academic communication is imbedded.

Learning theory and motivation then fall into place in explaining the dynamic process by which meanings are acquired and acted upon.

Measurement in education is introduced as a way of determining the meanings with which an individual comes to school, his learning context. It also helps us in assessing the new meanings which he acquires through the educational experience and his capacity for developing the more complex symbols. Life histories become important as an aid to understanding the socio-cultural meanings which a child brings to school. Sociograms and anecdotal records are presented as a means of gaining insight into their present social relations. The various achievement tests are treated in relation to the particular symbol ability or level that each assesses.

This approach to educational psychology which is built around communication theory as a theme allows us to view the whole child in action as in the individual psychological approach. In addition to this, more stress is placed upon the culture and the society in which he is developing. At the same time, by focusing throughout on the single theme of interaction which is mediated by symbols, a basis is indicated for logically integrating the contributions of the many fields upon which educational psychology draws. A final important advantage is that by concentrating on the actual symbols in real interaction processes, the work may be related to practical classroom situations at almost every step.

EVALUATION

Evaluating this course presented an interesting problem. If it were a matter of measuring student growth in subject matter alone, the procedure would have been clear. For that matter, a standardized short answer subject matter examination showed that the general knowledge of educational psychology gained was equal to that gained by students taught by the more traditional methods.

However, growth toward understanding how a whole individual learns in a whole environment could not be assessed so facilely.

This section of the article presents an attempt to use content analysis in measuring the attainment of this concept.[2]

A 1400-word short story was written describing two small boys from differing socio-economic strata and their experiences at home and at school. Here is a sample paragraph from the story:

> Mr. Walker (the teacher), who taught the sixth grade seemed always to be dressed in the same ill-fitting tweed suit. Secretly Matty wished Mr. Walker would like him. Yet, somehow, he was convinced that Mr. Walker's only purpose seemed to be in finding fault with him. Matty said that Mr. Walker was pretty mean. Matty's only contact with Stuart was to bitterly acuse him of being the teacher's pet. In fact, aside from accusing Stuart of that, he had little else to do with the boy and saw no reason to make friends with him. After all, they didn't do the same things.

The students were given the story to read at the beginning and end of the course. Each time they were asked to write seven hundred fifty words in answer to the question, "According to the story, what have the two boys learned? Cite specific evidence from the story."

The story was not discussed during the semester.

Each student essay was subjected to a content analysis in terms of cultural, social, psychological and academic learnings reported on the part of the boys. The first three are, of course, the traditional categories in which the social scientist analyzes the environment. The unit of analysis was any word, group of words or sentence which communicated a single meaning in terms of the four categories. Indices were set up for each content category. The student was scored one point for each index of learning that he recognized. It was assumed that the number of indices for all categories, reflected the extent to which the students realized that the boys were learning from their whole environment.

Students received credit for recognizing cultural indices if they wrote about the learning of values, symbols of status, life goals, way of life, etc. For example, a student would be credited with a cultural index if he said that the boys learned that the physical differences of their neighborhoods, as related in the story, symbolized differences in status. The social category included learnings about interpersonal relationships, social structure, rôles, authority situations, etc. A statement that the boys recognized the fact that they were members of different social classes would, for example, be an index. Credit for a psychological learning was given for such things as learning a self-concept, use of mechanisms of defense, etc. An illustration of this would be the statement that one of the boys learned to withdraw in the face of challenge. The academic category was restricted to the learning of the specific subject matter dealt with in the classroom situation which formed part of the story.

---

[2] Appreciation is expressed to Sanford Margolis for writing the story used in this evaluation and to Elizu Katz for acting as an independent judge in scoring the essays.

Students received credit only when the learnings which they reported were supported by citations from the content of the story.

An expert, who was trained in education and psychology, also performed this exercise in order to provide some notion as to the maximum number of indices that might theoretically be recognized for each category.

Table I summarizes the results of the content analysis.

A 't' test was performed to discover whether the differences in the number of indices observed at the beginning and end of the semester were significant. All differences proved significant at the .01 level except the academic which was little better than the .20 level. There may be two explanations for the low level of recognition of the academic indices. Firstly, the story did not emphasize the academic aspects of learning as evidenced by the small number in the report of the expert. Secondly, the previous training of these

Table I. Mean Number of Indices Recognized in Each Category at the Beginning and the End of the Semester and the Number Recognized by an Expert

N = 16 Students.

Category	Expert	Mean Student Score Before	Mean Student Score After	M_D
Cultural	60	7.6	18.1	10.5
Social	15	1.7	6.1	4.4
Psychological	43	6.9	18.7	11.8
Academic	10	0.8	1.5	0.7
Total	128	17.0	44.4	27.4

students had stressed the personality building aspects of education as part of the reaction against subject matter centered education.

Too much dependence cannot be placed on these figures however, because of the subjective element in scoring the student replies. An independent judge, who was trained in the content analysis of written communication, also scored the essays in order to provide a measure of reliability. The rank-order coefficient of correlation between the scores given by the author and those of the independent judge was .59.

In spite of its tentative nature, however, this evaluation lends weight to the hypothesis that thematic teaching of educational psychology in the context of communication theory can help students grow in their understanding that learning is the product of the interaction of a whole individual and his whole environment.

PREFERENCES IN COLORS AND ILLUSTRATIONS OF ELEMENTARY SCHOOL CHILDREN OF PUERTO RICO

ISMAEL RODRIGUEZ BOU and DAVID CRUZ LOPEZ

Superior Educational Council, University of Puerto Rico

In 1945 the Superior Educational Council of the University of Puerto Rico made a study of the reading facilities available for the elementary school children of the Island. The results obtained determined the undertaking of a series of research projects, one of which was a study of the preferences in colors and illustrations for books of the elementary schools, as expressed in the choices of the students themselves.

In Puerto Rico the textbooks in reading have constituted a serious educational problem not only because of the fact that we never had enough for our large elementary school enrollment, but also because the majority of them are inadequate in their contents, vocabulary and physical make-up.

A large number of the books used in the schools of the Island were originally written in English for children of the continental United States. They reflect the life, play activities, amusements, customs, and interests of these children. Psychologically, in many respects, they do not represent the Puerto Rican child. Many of them depict human types, landscape, flora, and fauna different from ours. The interior decoration and the appliances shown in the houses that appear in these books are usually different, too, from those of our environment.

Even the books written by Puerto Rican authors for Puerto Rican children do not meet the requirements, especially in regard to gradation of vocabulary and illustrations. The Spanish American texts used in some grades are still less well adapted to our use.

The Superior Educational Council started from the premise that our differences in culture and geographical medium from the continental United States justify differences in the illustrations of our school textbooks. We do not mean isolationism in this, especially today, when great emphasis is placed on human and international relations, but heretofore books have been illustrated generally from the point of view of the artists and writers without paying much attention to children's interests. The books used in our schools should reflect, of course, the life and geography of different parts of the world, but they should also take into consideration the realities of our native environment. Translations and adaptations of textbooks from other countries fulfilled their rôle in a time when they were the best that could be had. Henceforth our elementary school texts should interpret our ways of living and the influence which the external medium has exerted over us, without minimizing, of course, those themes and activities which are universally valid in the world of children.

A simple glance at the reading texts in our elementary schools reveals, besides the lack of adaptation to the environment already mentioned, many illustrations with poor color combinations, out of harmony with the light, colors, and tones of the tropical nature around us. Furthermore, many drawings have little artistic merit, and no uniformity whatsoever in regard to the number of illustrations used and the disposition of these illustrations on the pages.

These limitations made us think that a study of the preferences and dislikes of the Puerto Rican child in respect to colors, types of illustrations, and the position of these illustrations in relation to the text matter would undoubtedly help to better the quality of the books intended for our elementary schools.

AIMS

Accordingly we proposed to carry out research which would help us to determine:

1) The colors that our school children prefer for the covers and pictures of their reading books.

2) The types of illustrations they prefer for their books.

3) The part of the page where they wish the illustration to appear.

PREVIOUS RESEARCH

Neither the theme of the study nor the procedure employed has any claim to originality. In fact, in our research we came across fifteen or more studies carried on in the United States with similar aims and procedures. However, our study represents the first attempt to interpret with objectivity the preferences of the elementary school children of Puerto Rico in the items we wished to investigate. As far as we know, this is, too, the first study of its kind carried on in the Spanish-speaking countries.

SUBJECTS AND MATERIALS FOR THE EXPERIMENT

A representative sample of 2496 elementary school pupils was obtained. It included boys and girls of Grades II, IV and VI from both the urban and the rural zones. The different geographical, industrial, and agricultural areas of the Island were taken into consideration in the selection of the sample.

The materials for the experiment included:

A) For the color preferences:

1) Three cards 7″ x 9″, painted uniformly in a brilliant tone of equal intensity for each of the primary colors (red, yellow, blue), and another set of three similarly painted cards for each of the secondary colors (green, orange and violet). From these the students selected their favorite colors.

2) Thirty cards 7″ x 9″, for five tones of each of these six colors. The students were to select their favorite tone of the color already selected in (1). The five tones to be exhibited were: very light, light, normal, dark, and very dark.

3) Fifteen bichromatic combinations obtained with the six fundamental colors, in cards of size 14″ x 9″.

4) Twenty trichromatic combinations obtained with the six fundamental colors, in cards of size 18″ x 9″. In both (3) and (4) the pupils had to select their favorite combination.

B) For the different types of drawings:

Three drawings in ink, on white cardboard, 11″ x 8″. The subject of the three drawings was the same: the entrance door to a walled garden. One of the drawings was lineal, very schematic. Another was done in black and white areas. The third one was normal or realistic, a complete gradation of light and shadow, obtained with fine parallel and crossed lines and pen point work in black ink.

C) For the position of the illustrations:

We used ten white cards, each 7″ x 9″, containing a space in black to represent the actual size and the possible place of the illustration on the page. The spaces in black varied in size, from one-fourth page to a full page, and occupied different positions on the card.

Each card (for color, drawings, and position of the illustration) was assigned a key number to facilitate the handling of the material and the tabulation of the data.

In each school chosen for the experiment the students were picked at random and assembled in a room. The purpose of the experiment was explained to them and their personal data were recorded on a mimeographed questionnaire. The materials to be used were exposed in another room which had been previously equipped for that purpose. The students were sent to this room one at a time. They made their selections of color, type of drawing, and position of the illustration on the page, following the directions given by the persons in charge of the experiment.

The data were tabulated and analyzed.

GENERAL CONCLUSIONS

A) Monochromatic preferences:

1) Blue was the favorite color throughout all grades, ages (except 7), sexes and zones. Forty-two per cent of the 2496 students tested preferred it. This color was a stronger favorite with the older children than with the younger ones. Orange, on the other hand, was the least popular. Only 7.5 per cent of the tallies belong to it.

2) Red and yellow, in the order mentioned, came next in preference. These colors obtained 18.9 per cent and 14.4 per cent of the tallies, respectively.

3) Among the 108 seven-year old pupils red was the favorite color, with 27.8 per cent of these pupils voting for it. Blue, the next favorite for this age, obtained 25 per cent, while green, the least favored, obtained 4.6 per cent.

4) Primary colors were preferred to secondary colors. Green, violet and orange obtained 8.9 per cent, 8.3 per cent, and 7.5 per cent of the tallies, respectively.

5) The percentage of students who selected red, yellow and violet diminished with the advancement in grade. For red, the percentages were 19.8 per cent, 19.2 per cent and 17.7 per cent for Grades II, IV, and VI, respectively; for yellow, 17.1 per cent, 14.3 per cent and 11.8 per cent for the same grades in the same

order; for violet, 11.9 per cent, 10.0 per cent and 3.0 per cent in the same way.

6) The children from the rural zone showed a more decided preference for blue and red than the children from the urban zone (44.7 per cent for blue and 20.5 per cent for red in the rural zone, against 39.3 per cent for blue and 17.3 per cent for red in the urban zone). The reverse was true with yellow and orange (12.2 per cent for yellow and 5.4 per cent for orange in the rural zone, against 16.6 per cent for yellow and 9.6 per cent for orange in the urban zone).

B) Tones

1. The dark tone of each color was the favorite one.

2) The order of preference was the same in the three grades: dark, (35.7 per cent); normal, (23.4 per cent); light (21.5 per cent); very light (12.5 per cent); and very dark (6.9 per cent). The figures in parentheses indicate the per cent of 2496 students selecting each tone.

This same order prevailed throughout all ages (except for the youngest and the oldest children), sexes and zones. Very dark tones were, as a rule, rejected.

C) Bichromatic combinations

Blue-red was the bichromatic combination best liked, in general, by both boys and girls. It was, likewise, the first choice of the urban zone and the second choice of the rural zone, where blue-violet superseded it by 0.2 per cent. Nine-year olds and second-graders also preferred blue-violet. Orange-yellow, yellow-red, and red-green were the respective preferences of six-, seven- and eight-year olds. No general statement could be made regarding the bichromatic preferences by ages, except that the choices were rather erratic. The percentage range of preferences extended from 1.5 per cent (orange-violet, the least liked) to 14.6 per cent (blue-red, the most popular). The largest per cent difference between any two consecutive bichromatic combinations when arranged in descending order of preference, was 3.0, which came precisely between the best liked and the next favorite (blue-violet, 11.6 per cent).

D) Trichromatic combinations

Here the range of preference, expressed in per cent of the total number of students, extended from 1.5 (orange-green-violet, the least popular) to 10.3 (orange-blue-yellow, the best liked). When arranged in descending order of preference the differences between the consecutive percentages were small and rather smoothly distributed. The mode of the arrangement is 0.1 per cent, which occurred in five cases out of nineteen. The largest difference (2.2 per cent) came between the second and the third best liked combinations, 9.4 per cent and 7.2 per cent, for blue-yellow-red and blue-red-green, respectively.

The favorite trichromatic combination for the girls was blue-yellow-red; for the boys it was blue-yellow-orange. This combination was the girls' second choice, with a difference of only 0.5 per cent between it and their first choice. In both the urban and the rural zones, and in Grades II and IV, blue-yellow-orange was the preferred combination. In Grade VI blue-yellow-red was the first choice. The selections made by students of ages eleven, thirteen, fifteen, sixteen and seventeen (the oldest) departed from the general pattern of choices. No inclusive statement can be made for all of them.

TYPES OF ILLUSTRATIONS AND POSITION

The drawing with the most realistic approach to the theme (the complete gradation type) was by far the most popular in all grades, sexes and zones, and in almost all ages.

Those illustrations occupying either a full page or the upper half of the page had the approval of the students of all grades, sexes and zones, and nearly all ages.

CONCLUSIONS AND RECOMMENDATIONS

The analysis of the data gathered in the study reveals aesthetic concepts which possibly reflect the power of cultural influences on the Puerto Rican child, the sensibility awakened in him by the light and chromatic contrasts of a tropical environment, and the real and vicarious experiences he has lived. The choices made by the children may have been conditioned, too, by such factors as race, temperament, and physical and emotional state.

In general, neither age, sex, nor zone—urban or rural—influenced the elementary school children who participated in this experiment in their selection of colors, types of drawings or the place where the illustrations should appear on the page.

On the light of the findings of this research project the following recommendations are made with the aim of helping the persons who are interested in writing or buying textbooks for the children of Puerto Rico. School administrators and parents who frequently face the problem of selecting adequate books for children, may profit by them, too.

1) Covers for the books intended for the primary grades should be done in a rich tone of blue or red. The books assigned to the higher grades of the elementary school may have covers with blue, green or yellow as the main color.

Elementary school children will like, also, bichromatic covers in blue-red, blue-violet and blue-yellow. If three colors are preferred for the covers, those which will most likely meet the highest approval are blue-orange-yellow, blue-red-yellow, and blue-red-green.

2) Illustrations should be as life-like as possible, and in the favorite colors of children. These illustrations should occupy either a full page, or the upper half of the page. They should not divide or interrupt the printed part of the page.

Since the interests and preferences of children, including their interest in color and drawing, may be subject to change, they should be reëvaluated from time to time. These interests cannot by themselves be used as an absolute criterion in the preparation of reading textbooks. It is, of course, realized that other needs of children, the interests of society, and the opinion of experts should be taken in consideration as a counter-weight to the children's own subjective judgment.

PERCEPTUAL RETARDATION IN READING DISABILITY CASES

JAMES C. COLEMAN

The Clinical School

University of California at Los Angeles

The importance of inadequate reading readiness as an agent in reading disability has been well established. However there are a number of factors which go to make up the complex called reading readiness. Some of these, such as mental age, have been systematically studied and their rôle in reading disability pretty well delineated. Other factors, such as perception, have received little experimental attention although their importance is generally agreed upon in the literature (*2*, *3*). Harris (*3*), for example, observes: "Even if the eyes are normal, the child may have immature visual perception. Seeing a thing does not always mean noticing its details. Many young children pay attention only to the main characteristics of visual stimuli—the size, shape, and color—and ignore the details. When asked to match letters or words they make many errors, not because of faulty vision, but because they do not notice differences which are obvious to older children." (p. 29)

Implicit in such a statement is the assumption that perception has its own course and sequence of ontogenetic development. Gestalt psychologists have long maintained that this is so, and that such a development can best be described by the sequence: perception of the crude whole, differentiation of details, and integration of the differentiated parts into an articulated whole. In this connection Townsend (*6*) was able to point to some important trends in the maturational unfolding of the perceptual process. In an extended study of the factors which enter into the ability of copying, he came to the conclusion that copying is largely a function of perceptual ability, much more so than a function of motor skills. Between the mental ages of five to seven years, he found a rapid and regular increase in the reproduction of two aspects of figures; namely, components and essential form. By components he meant that all outstanding part aspects of a figure be depicted; for example, when copying a rectangle all four sides should be represented in the drawing. By essential form he meant that prominent shape characteristics of parts of figures be represented in the drawing; this implies, for example, that a small circle not be represented as a dot. It is evident, that this increase in components and essential form with mental age is one confirmation and exemplification of the Gestalt notion of differentiation.

The present study is an investigation of the gross development of visual perception in a group of reading disability cases. It is postulated that retardation in perceptual development is one important factor which enters into reading disability and that, consequently, this group of reading disability cases will be significantly inferior in the performance of perceptual tasks in comparison with a peer group which does not show any reading disability.

PROCEDURE

The subjects chosen for the present study represented a random sampling of forty male reading disability cases in attendance at the Clinical School of the University of California at Los Angeles. The subjects ranged in age from eight to forty-six years, of whom thirty-three were under thirteen years of age. All of the subjects were of average or better intelligence and free of disabling emotional or physical handicaps.

The non-verbal part of the Alpha Test of the Otis Quick-Scoring tests was chosen for the present investigation because it relies so heavily on perceptual factors. It consists of one hundred items of four pictures each. Three of these four pictures have one major aspect in common, while the fourth is different. The testee is instructed to draw a line through the one which is different. This test, therefore, requires sufficient perceptual discrimination to select a commonalty and to constitute a difference. While the more geometrical figures rely more heavily on perceptual discrimination (as, for example, in the four five-pointed stars, three of which have a dot on four points, while the fourth star has a dot on only three points—item 54, Form B), there are a number of items which depend to a greater extent on higher-order abstraction and concept formation. This is particularly true of some items which depict concrete objects. Item 62 (Form B), for example, depicts a clock, a ruler, a thermometer, and a pressing iron. The higher-order abstraction, 'measuring instrument', must be made before the item can be successfully solved. However, most of the concrete items do not appear to be as heavily loaded with higher-order concept formation as this one. Unfortunately, there are also a few items which require the ability to read figures and to do simple additions. However, most items require perceptual discrimination of what Townsend has called components and essential form. The over-all impression one gets is that the non-verbal part of the Otis Alpha is a fairly adequate instrument for the measurement of perceptual development, at least from the point of view of face validity.

By means of a table of age norms published in the Otis *Manual of Directions* (*5*) the point scores of the subjects were converted into age scores. The age score thus obtained, the mental age, is considered in this study to be representative of the perceptual age of the subject. If these reading disability cases are retarded in their perceptual development, then they should score on this test below the median for their age, and thus obtain a mental age below their chronological age. Such retardation gains in significance when we realize that the sample studied was selected from the upper half of the distribution of general intelligence. Furthermore, if we subtract the obtained mental age from the chronological age, we obtain an index of retardation. If, on the other hand, these pupils are not perceptually retarded, then the retardation index should be zero or negative.

In order to demonstrate that the results are not a function of a low level of intelligence, a perceptual-intellective lag index was computed. This index was obtained by subtracting the Otis Alpha MA from the MA obtained by other sources, specifically the Stanford-Binet, Wechsler-Bellevue (adult form), or WISC.

If perceptual retardation is found, then a positive perceptual-intellective index will demonstrate that it is not a function of the level of development of general intelligence, while a zero or negative perceptual-intellective lag index indicates that the perceptual retardation could be a function of the development of general intelligence.

RESULTS

The results obtained for the thirty-three children are shown in Table 1. They show that these children, as a group, are almost a year retarded in perceptual development when compared with

TABLE 1.—MEAN VALUES OF CHRONOLOGICAL AND MENTAL AGE, PERCEPTUAL RETARDATION, AND PERCEPTUAL-INTELLECTIVE LAG OF 33 CHILDREN WITH READING DISABILITY (COLUMNS 1, 3, 4, 5, 6 IN YEARS, COLUMN 2 IN POINT SCORES)

	1 CA	2 Otis Score	3 Otis MA	4 Retardation (MA-CA)	5 Other MA*	6 Perceptual Intellective lag (MA-OtisMA)
Mean	9.92	53.7	8.96*	.96	11.2	2.24
Standard error of mean				.31		.58
+				3.1		3.9
P				<1.0%		<1.0%

* Since a number of children had WISC or Wechsler IQ's, this mean is approximate, as the MA was estimated by $MA = \frac{IQ}{CA.}$

† This mean contains one estimated MA. Otis norms have been established only up to age 12–11, equivalent to a score of 72. The MA of one child who obtained a score of 74 was estimated at 14-0.

their age peer group. This mean retardation is significant beyond the .01 level of confidence.[1] That this result is not a function of the development of general intelligence is shown by the fact that the perceptual-intellective lag index is positive. This index reveals that the perceptual development of this group lags approximately two and a quarter years behind the development of general intelligence. This mean lag is also significant beyond the .01 level of confidence.

Of the thirty-three children, twenty-seven were perceptually retarded, the retardation ranging from one to forty-six months. On the other hand, six of these children were advanced beyond their age in their perceptual development, the advance ranging from one to fifty-four months.

While fifteen perceptually retarded children between the CA of 8-0 to 9-4 show a mean retardation of eleven months, twelve of the perceptually retarded children between the ages of 10-2 and 12-8 have a mean retardation of thirty-two months. This finding could indicate that perceptual retardation is cumulative.

---

[1] The means of Columns 1 and 3, as well as 3 and 5, were not directly compared because in such comparison age has to be partialled out. The method of the computation of a retardation and lag index avoids the necessity of computing a correlation coefficient.

Of the seven adults, ranging in age from thirteen to forty-six years, four were found to be perceptually retarded from nine to fifty-nine months, although their IQ's ranged from 98 to 106, with a median of 104. The other three were perceptually at least at par. They achieved scores higher than those for which age norms are available.

DISCUSSION

The fact that twenty of the forty subjects were retarded ten or more months in perceptual development would appear to be a very significant finding. It is reasonable to assume that such a marked degree of perceptual retardation would have a significant bearing on the reading disability of these subjects. However, this generalization cannot be extended to the entire group since a minority of subjects not only showed no retardation in perceptual development but considerable advance beyond their age. It would be futile to ascribe functional significance to perceptual factors in these reading disability cases—rather we must look for other factors.

A further trend which is indicated in the present data is that perceptual retardation is cumulative through the childhood age group; that is, the older the children get chronologically, the more retarded they become in relation to their ages. However, this trend does not appear to carry through into adulthood. Although there are too few adult cases to make any definite statements, it does appear that perceptual retardation plays a more significant rôle in the reading disability of children than in that of adults.

The question may be raised at this point as to what condition or conditions bring about a retardation in perceptual development in these subjects. Although formal evidence on this point is scanty, there are certain clinical impressions which may serve as helpful hints for further research: (a) Some children seem to be so absorbed in one specific interest, e.g., airplanes, chemistry, that they pay little attention to other normal objects of interest in their environment. Such an over-focussed interest in one area might well lead to an unequal differentiation of the perceptual field in which a small area becomes highly differentiated while other areas of the field remain relatively unstructured or undifferentiated. (b) Not infrequently children are observed who have been negatively conditioned to areas of perceptual discrimination. Here we usually find ambitious parents who put a great deal of pressure on their children to learn to read before they have reached a stage of reading readiness in which they can make the fine form discriminations essential to reading. The result is a state of anxious helplessness similar to what Goldstein calls 'catastrophic reaction' accompanied by avoidance reactions to such required perceptual discriminations. (c) In a number of these reading disability cases the concept of distinct perceptual types may be of value. There is considerable evidence for what we may term visual and haptic types (Lowenfeld *4*, Fernald *1*). The visual type uses his eyes as a mediator between himself and reality, while the haptic type relates to reality primarily through contact and kinesthetic sensations. When the emphasis in learning to read is exclusively on visual experiences without concomitant tactile and kinesthetic sensations, the haptic child may be at a distinct disadvantage. (d) Finally, there are a variety of emotional and personality factors which may enter into perceptual retardation. Children suffering from severe emotional conflicts and frustrations may be so chronically anxious that general learning and perceptual differentiation are adversely affected. Or such a child may focus his energies inward in self-preoccupation with his problems or in compensatory fantasies which interfere with his normal perceptual growth. It is interesting to note that Vorhaus (*7*) emphasizes such personality constellations in her study of personality patterns among non-readers.

Finally, this preliminary investigation would appear to have some implications for the symptomatic treatment of reading disability cases. It is evident that a majority of these cases can profit from remedial instruction directed toward the better differentiation and integration of perceptual experiences. It is also evident that there is a substantial minority of cases whose reading disability will not be materially improved by this type of instruction.

SUMMARY

This study dealt with the hypothesized relationship between perceptual retardation and reading disability in subjects with average or above average intelligence. It was found that in this group of reading disability cases:

1) A majority of these subjects were found to be markedly retarded in perceptual development.

2) Perceptual development lagged significantly behind the development of general intelligence in a majority of subjects.

3) A minority of subjects were average or above in perceptual development.

4) Retardation in perceptual differentiation was cumulative with age.

5) Perceptual retardation was a significant factor in reading disability.

The possible value of training in perceptual differentiation as symptomatic treatment for reading disability is suggested.

REFERENCES

1) Grace M. Fernald. *Remedial Techniques in Basic School Subjects*. New York: McGraw-Hill Book Co., 1943.

2) Arthur I. Gates. *The Improvement of Reading*. New York: The Macmillan Co., 1950.

3) A. J. Harris. *How To Increase Reading Ability*. New York: Longmans, Green, 1949.

4) Viktor Lowenfeld. *The Nature of Creative Activity*. New York: Harcourt, Brace, 1939.

5) A. S. Otis. *Manual of Directions for Alpha Test, Otis Quick-scoring Mental Ability Tests*. New York: World Book Co., 1939.

6) E. A. Townsend. *A Study of Copying Ability in Children, Genetic Psychol. Monogr.* 43, 1951, 1–49.

7) P. G. Vorhaus. "Rorschach configurations associated with reading disability." *J. Projective Techn.* 16, 1952, 1–19.